

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2003年 1月20日

出 願 番 号

Application Number:

特願2003-011594

[ST.10/C]:

[JP2003-011594]

出 願 人

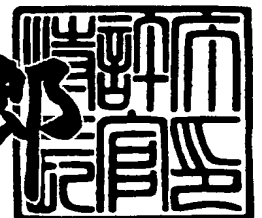
Applicant(s):

株式会社日立製作所

2003年 6月13日

特 許 庁 長 官
Commissioner,
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3046498

【書類名】 特許願

【整理番号】 HI020926

【提出日】 平成15年 1月20日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/06

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内

【氏名】 ▲高▼田 豊

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内

【氏名】 中山 信一

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内

【氏名】 小笠原 裕

【発明者】

【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所ソフトウェア事業部内

【氏名】 志川 甚一

【発明者】

【住所又は居所】 神奈川県横浜市戸塚区戸塚町 5 0 3 0 番地 株式会社日立製作所ソフトウェア事業部内

【氏名】 雑賀 信之

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100071283

【弁理士】

【氏名又は名称】 一色 健輔

【選任した代理人】

【識別番号】 100084906

【弁理士】

【氏名又は名称】 原島 典孝

【選任した代理人】

【識別番号】 100098523

【弁理士】

【氏名又は名称】 黒川 恵

【選任した代理人】

【識別番号】 100112748

【弁理士】

【氏名又は名称】 吉田 浩二

【選任した代理人】

【識別番号】 100110009

【弁理士】

【氏名又は名称】 青木 康

【手数料の表示】

【予納台帳番号】 011785

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【ブルーフの要否】 要

【書類名】 明細書

【発明の名称】 記憶デバイス制御装置の制御方法、及び記憶デバイス制御装置

【特許請求の範囲】

【請求項 1】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I/O 要求を出力する I/O プロセッサとが形成された回路基板を有する複数のチャネル制御部と、

前記 I/O プロセッサから送信される前記 I/O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

を含んで構成され、

前記記憶デバイスにより提供される記憶領域をその記憶領域に論理的に設定される記憶領域である論理ボリュームを単位として管理する記憶デバイス制御装置の制御方法であって、

前記ディスク制御部が、第一の論理ボリュームにデータが書き込まれた場合にそのデータの複製を第二の論理ボリュームにも記憶するために前記データを第二の論理ボリュームにも書き込む複製管理処理を実行するステップ

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項 2】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I/O 要求を出力する I/O プロセッサとが形成された回路基板を有する複数のチャネル制御部と、

前記 I/O プロセッサから送信される前記 I/O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

を含んで構成され、

前記記憶デバイスにより提供される記憶領域をその記憶領域に論理的に設定される記憶領域である論理ボリュームを単位として管理する記憶デバイス制御装置の制御方法であって、

前記ディスク制御部が、第一の論理ボリュームにデータが書き込まれた場合にそのデータの複製を第二の論理ボリュームにも記憶するために前記データを第二

の論理ボリュームにも書き込む複製管理処理を実行するステップと、

前記チャネル制御部が、前記情報処理装置から送信される、前記複製の対象とするデータをファイル単位またはディレクトリ単位で指定する情報を受信するステップと、

前記チャネル制御部が、前記情報を受信してこれに指定されるファイルやディレクトリに対応するデータを特定し、そのデータが前記第一の論理ボリュームに記憶されるように制御するステップと、

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項 3】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有する複数のチャネル制御部と、

前記 I / O プロセッサから送信される前記 I / O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

を含んで構成され、

前記記憶デバイスにより提供される記憶領域をその記憶領域に論理的に設定される記憶領域である論理ボリュームを単位として管理する記憶デバイス制御装置の制御方法であって、

前記ディスク制御部が、第一の論理ボリュームにデータが書き込まれた場合にそのデータの複製を第二の論理ボリュームにも記憶するために前記データを第二の論理ボリュームにも書き込む複製管理処理を実行するステップと、

前記チャネル制御部が、前記情報処理装置から送信される、前記複製の対象とするデータをファイル単位またはディレクトリ単位で指定する第一の情報を受信するステップと、

前記チャネル制御部が、前記第一の情報を受信してこれに指定されるファイルやディレクトリに対応するデータを特定し、そのデータが前記第一の論理ボリュームに記憶されるように制御するステップと、

前記チャネル制御部が、前記情報処理装置から送信される、前記複製管理処理の停止を指示する第二の情報を受信するステップと、

前記チャンネル制御部が、前記第二の情報を受信した場合にその旨を前記ディスク制御部に通知するステップと、

前記ディスク制御部が、前記通知を受信した場合に前記複製管理処理を停止するステップと、

前記チャンネル制御部が、前記第二の情報を受信した場合に、前記第一の情報に指定されるファイルやディレクトリに対応するデータについての前記記憶デバイスへの書き込みにより複数の前記第一の論理ボリュームへの書き込みが生じる場合には、前記複製管理処理が停止するのを待ってから前記第一の論理ボリュームへの書き込みを開始するステップと、

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項4】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャンネル制御部と、

前記I/Oプロセッサから送信される前記I/O要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

を含んで構成され、

前記記憶デバイスにより提供される記憶領域をその記憶領域に論理的に設定される記憶領域である論理ボリュームを単位として管理する記憶デバイス制御装置の制御方法であって、

前記ディスク制御部が、第一の論理ボリュームにデータが書き込まれた場合にそのデータの複製を他の記憶デバイス制御装置により提供される第二の論理ボリュームにおいても記憶するために前記データを前記他の記憶デバイス制御装置に送信する処理を実行するステップと、

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項5】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャンネル制御部と、

前記 I / O プロセッサから送信される前記 I / O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

を含んで構成され、

前記記憶デバイスにより提供される記憶領域をその記憶領域に論理的に設定される記憶領域である論理ボリュームを単位として管理する記憶デバイス制御装置の制御方法であって、

前記ディスク制御部が、第一の論理ボリュームにデータが書き込まれた場合にそのデータの複製を他の記憶デバイス制御装置により提供される第二の論理ボリュームにおいても記憶するために前記データを前記他の記憶デバイス制御装置に送信する処理を実行するステップと、

前記チャネル制御部が、前記情報処理装置から送信される、前記複製の対象となるデータをファイル単位またはディレクトリ単位で指定する情報を受信するステップと、

前記チャネル制御部が、前記情報を受信してこれに指定されるファイルやディレクトリに対応するデータを特定し、そのデータが前記第一の論理ボリュームに記憶されるように制御するステップと、

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項 6】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有する複数のチャネル制御部と、

前記 I / O プロセッサから送信される前記 I / O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

を含んで構成され、

前記記憶デバイスにより提供される記憶領域をその記憶領域に論理的に設定される記憶領域である論理ボリュームを単位として管理する記憶デバイス制御装置であって、

前記ディスク制御部が、第一の論理ボリュームにデータが書き込まれた場合にそのデータの複製を第二の論理ボリュームにも記憶するために前記データを第二

の論理ボリュームにも書き込む複製管理処理を実行する手段
を備えることを特徴とする記憶デバイス制御装置。

【請求項 7】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有する複数のチャネル制御部と、

前記 I / O プロセッサから送信される前記 I / O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、
を含んで構成され、

前記記憶デバイスにより提供される記憶領域をその記憶領域に論理的に設定される記憶領域である論理ボリュームを単位として管理する記憶デバイス制御装置であって、

前記ディスク制御部が、第一の論理ボリュームにデータが書き込まれた場合にそのデータの複製を第二の論理ボリュームにも記憶するために前記データを第二の論理ボリュームにも書き込む複製管理処理を実行する手段と、

前記チャネル制御部が、前記情報処理装置から送信される、前記複製の対象とするデータをファイル単位またはディレクトリ単位で指定する情報を受信する手段と、

前記チャネル制御部が、前記情報を受信してこれに指定されるファイルやディレクトリに対応するデータを特定し、そのデータが前記第一の論理ボリュームに記憶されるように制御する手段と、

を備えることを特徴とする記憶デバイス制御装置。

【請求項 8】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有する複数のチャネル制御部と、

前記 I / O プロセッサから送信される前記 I / O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

を含んで構成され、

前記記憶デバイスにより提供される記憶領域をその記憶領域に論理的に設定される記憶領域である論理ボリュームを単位として管理する記憶デバイス制御装置であって、

前記ディスク制御部が、第一の論理ボリュームにデータが書き込まれた場合にそのデータの複製を第二の論理ボリュームにも記憶するために前記データを第二の論理ボリュームにも書き込む複製管理処理を実行する手段と、前記情報処理装置から送信される、前記複製の対象とするデータをファイル単位またはディレクトリ単位で指定する第一の情報を受信する手段と、前記第一の情報を受信してこれに指定されるファイルやディレクトリに対応するデータを特定し、そのデータが前記第一の論理ボリュームに記憶されるように制御する手段と、前記情報処理装置から送信される、前記複製管理処理の停止を指示する第二の情報を受信する手段と、前記第二の情報を受信した場合にその旨を前記ディスク制御部に通知する手段とを備え、

前記ディスク制御部が、前記通知を受信した場合に前記複製管理処理を停止する手段を備え、

前記チャネル制御部が、前記第二の情報を受信した場合に、前記第一の情報に指定されるファイルやディレクトリに対応するデータについての前記記憶デバイスへの書き込みにより複数の前記第一の論理ボリュームへの書き込みが生じる場合には、前記複製管理処理が停止するのを待ってから前記第一の論理ボリュームへの書き込みを開始する手段

を備えることを特徴とする記憶デバイス制御装置。

【請求項9】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャネル制御部と、

前記I/Oプロセッサから送信される前記I/O要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

を含んで構成され、

前記記憶デバイスにより提供される記憶領域をその記憶領域に論理的に設定さ

れる記憶領域である論理ボリュームを単位として管理する記憶デバイス制御装置であって、

前記ディスク制御部が、第一の論理ボリュームにデータが書き込まれた場合にそのデータの複製を他の記憶デバイス制御装置により提供される第二の論理ボリュームにおいても記憶するために前記データを前記他の記憶デバイス制御装置に送信する処理を実行する手段

を備えることを特徴とする記憶デバイス制御装置。

【請求項 1 0】 情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有する複数のチャネル制御部と、

前記 I / O プロセッサから送信される前記 I / O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

を含んで構成され、

前記記憶デバイスにより提供される記憶領域をその記憶領域に論理的に設定される記憶領域である論理ボリュームを単位として管理する記憶デバイス制御装置であって、

前記ディスク制御部が、第一の論理ボリュームにデータが書き込まれた場合にそのデータの複製を他の記憶デバイス制御装置により提供される第二の論理ボリュームにおいても記憶するために前記データを前記他の記憶デバイス制御装置に送信する処理を実行する手段と、

前記チャネル制御部が、前記情報処理装置から送信される、前記複製の対象となるデータをファイル単位またはディレクトリ単位で指定する情報を受信する手段と、前記情報を受信してこれに指定されるファイルやディレクトリに対応するデータを特定し、そのデータが前記第一の論理ボリュームに記憶されるように制御する手段と

を備えることを特徴とする記憶デバイス制御装置。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は、記憶デバイス制御装置の制御方法、及び記憶デバイス制御装置に関する。

【0002】

【従来の技術】

近年コンピュータシステムで取り扱われるデータ量が急激に増加している。このようなデータを管理するためのストレージシステムとして、最近ではミッドレンジクラスやエンタープライズクラスと呼ばれるような、巨大な記憶資源を提供する R A I D (Redundant Arrays of Inexpensive Disks) 方式で管理された大規模なストレージシステムが注目されている。また、かかる膨大なデータを効率よく利用し管理するために、ディスクアレイ装置等のストレージシステムと情報処理装置とを専用のネットワーク (Storage Area Network、以下 S A N と記す) で接続し、ストレージシステムへの高速かつ大量なアクセスを実現する技術が開発されている。

一方、ストレージシステムと情報処理装置とを T C P / I P (Transmission Control Protocol/Internet Protocol) プロトコル等を用いたネットワークで相互に接続し、情報処理装置からのファイルレベルでのアクセスを実現する N A S (Network Attached Storage) と呼ばれるストレージシステムが開発されている。

【0003】

【特許文献1】

特開 2 0 0 2 - 3 5 1 7 0 3 号公報

【0004】

【発明が解決しようとする課題】

しかしながら従来の N A S は、T C P / I P 通信機能及びファイルシステム機能を持たない記憶装置に、T C P / I P 通信機能及びファイルシステム機能を持った情報処理装置を付加することにより実現されていた。そのため、上記付加される情報処理装置の設置スペースが必要であった。また上記情報処理装置と記憶装置との間は、高速に通信を行う必要性から S A N で接続されていることが多く

、そのための通信制御機器や通信制御機能を備える必要もあった。

本発明は上記課題を鑑みてなされたものであり、記憶デバイス制御装置の制御方法、及び記憶デバイス制御装置を提供することを主たる目的とする。

【 0 0 0 5 】

【課題を解決するための手段】

上記目的を達成するための本発明の主たる発明は、

情報処理装置から送信されるファイル単位でのデータ入出力要求をネットワークを通じて受信するファイルアクセス処理部と、記憶デバイスに対する前記データ入出力要求に対応する I / O 要求を出力する I / O プロセッサとが形成された回路基板を有する複数のチャネル制御部と、

前記 I / O プロセッサから送信される前記 I / O 要求に応じて前記記憶デバイスに対するデータ入出力を実行するディスク制御部と、

を含んで構成され、

前記記憶デバイスにより提供される記憶領域をその記憶領域に論理的に設定される記憶領域である論理ボリュームを単位として管理する記憶デバイス制御装置の制御方法であって、

前記ディスク制御部が、第一の論理ボリュームにデータが書き込まれた場合にそのデータの複製を第二の論理ボリュームにも記憶するために前記データを第二の論理ボリュームにも書き込む複製管理処理を実行するステップ

を備えることとする。

【 0 0 0 6 】

なお、前記情報処理装置とは、前記構成の前記記憶デバイス制御装置を備えて構成されるストレージシステムに LAN や SAN を介してアクセスする、例えば、パーソナルコンピュータやメインフレームコンピュータである。ファイルアクセス処理部の機能は CPU 上で実行されるオペレーティングシステムおよびこのオペレーティングシステム上で動作する例えば NFS (Network File System) 等のソフトウェアによって提供される。記憶デバイスは例えばハードディスク装置などのディスクドライブである。I / O プロセッサは例えばファイルアクセス処理部のハードウェア要素である前記 CPU とは独立した IC (Integrated Cir

cuit) をハードウェア要素とし、ファイルアクセス処理部とディスク制御部との間の通信を制御する。ディスク制御部は、記憶デバイスに対してデータの書き込みや読み出しを行う。

その他、本願が開示する課題、及びその解決方法は、発明の実施の形態の欄、及び図面により明らかにされる。

【 0 0 0 7 】

【発明の実施の形態】

===全体構成例===

ストレージシステム 6 0 0 は、記憶デバイス制御装置 1 0 0 と記憶デバイス 3 0 0 とを備えている。記憶デバイス制御装置 1 0 0 は、情報処理装置 2 0 0 から受信したコマンドに従って記憶デバイス 3 0 0 に対する制御を行う。例えば情報処理装置 2 0 0 からデータの入出力要求を受信して、記憶デバイス 3 0 0 に記憶されているデータの入出力のための処理を行う。データは、記憶デバイス 3 0 0 が備えるディスクドライブにより提供される物理的な記憶領域上に論理的に設定される記憶領域である論理ボリューム (Logical Unit) (以下、L U と記す) に記憶されている。また記憶デバイス制御装置 1 0 0 は、情報処理装置 2 0 0 との間で、ストレージシステム 6 0 0 を管理するための各種コマンドの授受も行う。

【 0 0 0 8 】

情報処理装置 2 0 0 は C P U (Central Processing Unit) やメモリを備えたコンピュータである。情報処理装置 2 0 0 が備える C P U により各種プログラムが実行されることにより様々な機能が実現される。情報処理装置 2 0 0 は、例えばパーソナルコンピュータやワークステーションであることもあるし、メインフレームコンピュータであることもある。

【 0 0 0 9 】

図 1 において、情報処理装置 1 乃至 3 (2 0 0) は、L A N (Local Area Network) 4 0 0 を介して記憶デバイス制御装置 1 0 0 と接続されている。L A N 4 0 0 は、インターネットとすることもできるし、専用のネットワークとすることもできる。L A N 4 0 0 を介して行われる情報処理装置 1 乃至 3 (2 0 0) と記憶デバイス制御装置 1 0 0 との間の通信は、例えば T C P / I P プロトコルに従

って行われる。情報処理装置1乃至3(200)からは、ストレージシステム600に対して、ファイル名指定によるデータアクセス要求(ファイル単位でのデータ入出力要求。以下、ファイルアクセス要求と記す)が送信される。

【0010】

LAN400にはバックアップデバイス910が接続されている。バックアップデバイス910は具体的にはMOやCD-R、DVD-RAMなどのディスク系デバイス、DATテープ、カセットテープ、オープンテープ、カートリッジテープなどのテープ系デバイスである。バックアップデバイス910は、LAN400を介して記憶デバイス制御装置100との間で通信を行うことにより、記憶デバイス300に記憶されているデータのバックアップデータを記憶する。またバックアップデバイス910は情報処理装置1(200)と接続されるようにすることもできる。この場合は情報処理装置1(200)を介して記憶デバイス300に記憶されているデータのバックアップデータを取得するようにする。

【0011】

記憶デバイス制御装置100は、チャンネル制御部1乃至4(110)を備える。記憶デバイス制御装置100は、チャンネル制御部1乃至4(110)によりLAN400を介して情報処理装置1乃至3(200)やバックアップデバイス910との間で通信を行う。チャンネル制御部1乃至4(110)は、情報処理装置1乃至3(200)からのファイルアクセス要求を個々に受け付ける。すなわち、チャンネル制御部1乃至4(110)には、個々にLAN400上のネットワークアドレス(例えば、IPアドレス)が割り当てられていてそれぞれが個別にNASとして振る舞い、個々のNASがあたかも独立したNASが存在しているかのようにNASのサービスを情報処理装置1乃至3(200)に提供することができる。以下、チャンネル制御部1乃至4(110)をCHNと記す。このように1台のストレージシステム600に個別にNASとしてのサービスを提供するチャンネル制御部1乃至4(110)を備えるように構成したことで、従来、独立したコンピュータで個別に運用されていたNASサーバが一台のストレージシステム600に集約されて運用される。そして、この構成によってストレージシステム600の統括的な管理が可能となり、各種設定・制御や障害管理、バージョン

管理といった保守業務の効率化が図られる。

【0012】

なお、本実施の形態に係る記憶デバイス制御装置100のチャネル制御部1乃至4（110）の機能は、後述するように、一体的にユニット化された回路基板上に形成されたハードウェア及びこのハードウェアにより実行されるオペレーティングシステム（以下、OSと記す）やこのOS上で動作するアプリケーションプログラムなどのソフトウェアにより実現される。このように本実施例のストレージシステム600では、従来ハードウェアの一部として実装されてきた機能が主としてソフトウェアによって実現されている。このため、本実施例のストレージシステム600では柔軟性に富んだシステム運用が可能となり、多様で変化の激しいユーザニーズに対応したきめ細かなサービスを提供することが可能となる。

【0013】

情報処理装置3乃至4（200）はSAN（Storage Area Network）500を介して記憶デバイス制御装置100と接続されている。SAN500は、記憶デバイス300が提供する記憶領域におけるデータの管理単位であるブロックを単位として情報処理装置3乃至4（200）との間でデータの授受を行うためのネットワークである。SAN500を介して行われる情報処理装置3乃至4（200）と記憶デバイス制御装置100との間の通信は、一般にファイバチャネルプロトコルに従って行われる。情報処理装置3乃至4（200）からは、ストレージシステム600に対して、ファイバチャネルプロトコルに従ってブロック単位のデータアクセス要求（以下、ブロックアクセス要求と記す）が送信される。

【0014】

SAN500にはSAN対応のバックアップデバイス900が接続されている。SAN対応バックアップデバイス900は、SAN500を介して記憶デバイス制御装置100との間で通信を行うことにより、記憶デバイス300に記憶されているデータのバックアップデータを記憶する。

【0015】

記憶デバイス制御装置100は、チャネル制御部5乃至6（110）を備える

。記憶デバイス制御装置100は、チャンネル制御部5乃至6(110)によりSAN500を介して情報処理装置3乃至4(200)及びSAN対応バックアップデバイス900との間の通信を行う。以下、チャンネル制御部5乃至6(110)をCHFと記す。

【0016】

また情報処理装置5(200)は、LAN400やSAN500等のネットワークを介さずに記憶デバイス制御装置100と接続されている。情報処理装置5(200)としては例えばメインフレームコンピュータとすることができる。情報処理装置5(200)と記憶デバイス制御装置100との間の通信は、例えばFICON(Fibre Connection)(登録商標)やESCON(Enterprise System Connection)(登録商標)、ACONARC(Advanced Connection Architecture)(登録商標)、FIBARC(Fibre Connection Architecture)(登録商標)などの通信プロトコルに従って行われる。情報処理装置5(200)からは、ストレージシステム600に対して、これらの通信プロトコルに従ってブロックアクセス要求が送信される。

記憶デバイス制御装置100は、チャンネル制御部7乃至8(110)により情報処理装置5(200)との間で通信を行う。以下、チャンネル制御部7乃至8(110)をCHAと記す。

【0017】

SAN500にはストレージシステム600の設置場所(プライマリサイト)とは遠隔した場所(セカンダリサイト)に設置される他のストレージシステム610が接続している。ストレージシステム610は、後述するリモートコピー(遠隔複製)又はレプリケーションの機能におけるデータの複製先の装置として利用される。なお、ストレージシステム610はSAN500以外にもATMなどの通信回線によりストレージシステム600に接続していることもある。この場合には例えばチャンネル制御部110として上記通信回線を利用するためのインターフェース(チャンネルエクステンダ)を備えるチャンネル制御部110が採用される。

【0018】

本実施例によれば、ストレージシステム600内にCHN110、CHF110、CHA110を混在させて装着させることにより、異種ネットワークに接続されるストレージシステムを実現できる。具体的には、ストレージシステム600は、CHN110を用いてLAN140に接続し、かつCHF110を用いてSAN500に接続するという、SAN-NAS統合ストレージシステムである。

【0019】

==記憶デバイス==

記憶デバイス300は多数のディスクドライブ（物理ディスク）を備えており情報処理装置200に対して記憶領域を提供する。データは、ディスクドライブにより提供される物理的な記憶領域上に論理的に設定される記憶領域であるLUに記憶されている。ディスクドライブとしては、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等様々なものを用いることができる。

【0020】

なお、記憶デバイス300は例えば複数のディスクドライブによりディスクアレイを構成するようにすることもできる。この場合、情報処理装置200に対して提供される記憶領域は、RAIDにより管理された複数のディスクドライブにより提供されるようにすることもできる。

記憶デバイス制御装置100と記憶デバイス300との間は図1のように直接に接続される形態とすることもできるし、ネットワークを介して接続するようにすることもできる。さらに記憶デバイス300は記憶デバイス制御装置100と一体として構成されることもできる。

【0021】

記憶デバイス300に設定されるLUには、情報処理装置200からアクセス可能なユーザLUや、チャネル制御部110の制御のために使用されるシステムLU等がある。システムLUにはCHN110で実行されるオペレーティングシステムも格納される。また各LUにはチャネル制御部110が対応付けられている。これによりチャネル制御部110毎にアクセス可能なLUが割り当てられている。また上記対応付けは、複数のチャネル制御部110で一つのLUを共有す

るようにすることもできる。なお以下において、ユーザLUやシステムLUをユーザディスク、システムディスク等とも記す。また、複数のチャネル制御部110で共有されるLUを共有LUあるいは共有ディスクと記す。

【0022】

===記憶デバイス制御装置===

記憶デバイス制御装置100はチャネル制御部110、共有メモリ120、キャッシュメモリ130、ディスク制御部140、管理端末160、接続部150を備える。

【0023】

チャネル制御部110は情報処理装置200との間で通信を行うための通信インタフェースを備え、情報処理装置200との間でデータ入出力コマンド等を授受する機能を備える。例えばCHN110は情報処理装置1乃至3(200)からのファイルアクセス要求を受け付ける。そしてファイルの記憶アドレスやデータ長等を求めて、ファイルアクセス要求に対応するI/O要求を出力することにより、記憶デバイス300へのアクセスを行う。これによりストレージシステム600はNASとしてのサービスを情報処理装置1乃至3(200)に提供することができる。なおI/O要求にはデータの先頭アドレス、データ長、読み出し又は書き込み等のアクセスの種別が含まれている。またデータの書き込みの場合にはI/O要求には書き込みデータが含まれているようにすることもできる。I/O要求の出力は、後述するI/Oプロセッサ119により行われる。またCHF110は情報処理装置3乃至4(200)からのファイバチャネルプロトコルに従ったブロックアクセス要求を受け付ける。これによりストレージシステム600は高速アクセス可能なデータ記憶サービスを情報処理装置3乃至4(200)に対して提供することができる。またCHA110は情報処理装置5(200)からのFICONやESCON、ACONARC、FIBARC等のプロトコルに従ったブロックアクセス要求を受け付ける。これによりストレージシステム600は情報処理装置5(200)のようなメインフレームコンピュータに対してもデータ記憶サービスを提供することができる。

【0024】

各チャネル制御部110は管理端末160と共に内部LAN151で接続されている。これによりチャネル制御部110に実行させるマイクロプログラム等を管理端末160から送信しインストールすることが可能となっている。チャネル制御部110の構成については後述する。

【0025】

接続部150はチャネル制御部110、共有メモリ120、キャッシュメモリ130、ディスク制御部140を相互に接続する。チャネル制御部110、共有メモリ120、キャッシュメモリ130、ディスク制御部140間でのデータやコマンドの授受は接続部150を介することにより行われる。接続部150は例えば高速スイッチングによりデータ伝送を行う超高速クロスバススイッチなどの高速バスである。チャネル制御部110同士が高速バスで接続されていることで、個々のコンピュータ上で動作するNASサーバがLANを通じて接続する従来の構成に比べてチャネル制御部110間の通信パフォーマンスが大幅に向上する。またこれにより高速なファイル共有機能や高速フェイルオーバなども可能となる。

【0026】

共有メモリ120及びキャッシュメモリ130は、チャネル制御部110、ディスク制御部140により共有される記憶メモリである。共有メモリ120は主に制御情報やコマンド等を記憶するために利用されるのに対し、キャッシュメモリ130は主にデータを記憶するために利用される。

【0027】

例えば、あるチャネル制御部110が情報処理装置200から受信したデータ入出力コマンドが書き込みコマンドであった場合には、当該チャネル制御部110は書き込みコマンドを共有メモリ120に書き込むと共に、情報処理装置200から受信した書き込みデータをキャッシュメモリ130に書き込む。一方、ディスク制御部140は共有メモリ120を監視しており、共有メモリ120に書き込みコマンドが書き込まれたことを検出すると、当該コマンドに従ってキャッシュメモリ130から書き込みデータを読み出して記憶デバイス300に書き込む。

【0028】

またあるチャネル制御部110が情報処理装置200から受信したデータ入出力コマンドが読み出しコマンドであった場合には、当該チャネル制御部110は読み出しコマンドを共有メモリ120に書き込むと共に、読み出し対象となるデータがキャッシュメモリ130に存在するかどうかを調べる。ここでキャッシュメモリ130に存在すれば、チャネル制御部110はそのデータを情報処理装置200に送信する。一方、読みだし対象となるデータがキャッシュメモリ130に存在しない場合には、共有メモリ120を監視することにより読み出しコマンドが共有メモリ120に書き込まれたことを検出したディスク制御部140が、記憶デバイス300から読みだし対象となるデータを読み出してこれをキャッシュメモリ130に書き込むと共に、その旨を共有メモリ120に書き込む。そして、チャネル制御部110は共有メモリ120を監視することにより読みだし対象となるデータがキャッシュメモリ130に書き込まれたことを検出すると、そのデータを情報処理装置200に送信する。

【0029】

なお、このようにチャネル制御部110からディスク制御部140に対するデータの書き込みや読み出しの指示を共有メモリ120を介在させて間接に行う構成の他、例えばチャネル制御部110からディスク制御部140に対してデータの書き込みや読み出しの指示を共有メモリ120を介さずに直接に行う構成とすることもできる。

【0030】

ディスク制御部140は記憶デバイス300の制御を行う。例えば上述のように、チャネル制御部110が情報処理装置200から受信したデータ書き込みコマンドに従って記憶デバイス300へデータの書き込みを行う。また、チャネル制御部110により送信された論理アドレス指定によるLUへのデータアクセス要求を、物理アドレス指定による物理ディスクへのデータアクセス要求に変換する。記憶デバイス300における物理ディスクがRAIDにより管理されている場合には、RAID構成に従ったデータのアクセスを行う。またディスク制御部140は、記憶デバイス300に記憶されたデータの複製管理の制御やバックア

ップ制御を行う。さらにディスク制御部 1 4 0 は、災害発生時のデータ消失防止（ディザスタリカバリ）などを目的としてプライマリサイトのストレージシステム 6 0 0 のデータの複製をセカンダリサイトに設置された他のストレージシステム 6 1 0 にも記憶する制御（レプリケーション機能、又はリモートコピー機能）なども行う。

【 0 0 3 1 】

各ディスク制御部 1 4 0 は管理端末 1 6 0 と共に内部 LAN 1 5 1 で接続されており、相互に通信を行うことが可能である。これにより、ディスク制御部 1 4 0 に実行させるマイクロプログラム等を管理端末 1 6 0 から送信しインストールすることが可能となっている。ディスク制御部 1 4 0 の構成については後述する。本実施例においては、共有メモリ 1 2 0 及びキャッシュメモリ 1 3 0 がチャンネル制御部 1 1 0 及びディスク制御部 1 4 0 に対して独立に設けられていることについて記載したが、本実施例はこの場合に限られるものでなく、共有メモリ 1 2 0 又はキャッシュメモリ 1 3 0 がチャンネル制御部 1 1 0 及びディスク制御部 1 4 0 の各々に分散されて設けられることも好ましい。この場合、接続部 1 5 0 は、分散された共有メモリ又はキャッシュメモリを有するチャンネル制御部 1 1 0 及びディスク制御部 1 4 0 を相互に接続させることになる。

【 0 0 3 2 】

===管理端末===

管理端末 1 6 0 はストレージシステム 6 0 0 を保守・管理するためのコンピュータである。管理端末 1 6 0 を操作することにより、例えば記憶デバイス 3 0 0 内の物理ディスク構成の設定や、LUの設定、チャンネル制御部 1 1 0 において実行されるマイクロプログラムのインストール等を行うことができる。ここで、記憶デバイス 3 0 0 内の物理ディスク構成の設定としては、例えば物理ディスクの増設や減設、RAID構成の変更（RAID 1 から RAID 5 への変更等）等を行うことができる。さらに管理端末 1 6 0 からは、ストレージシステム 6 0 0 の動作状態の確認や故障部位の特定、チャンネル制御部 1 1 0 で実行されるオペレーティングシステムのインストール等の作業を行うこともできる。また管理端末 1 6 0 は LAN や電話回線等で外部保守センタと接続されており、管理端末 1 6 0

を利用してストレージシステム600の障害監視を行ったり、障害が発生した場合に迅速に対応することも可能である。障害の発生は例えばOSやアプリケーションプログラム、ドライバソフトウェアなどから通知される。この通知はHTTPプロトコルやSNMP (Simple Network Management Protocol)、電子メールなどにより行われる。これらの設定や制御は、管理端末160で動作するWebサーバが提供するWebページをユーザインタフェースとしてオペレータなどにより行われる。オペレータ等は、管理端末160を操作して障害監視する対象や内容の設定、障害通知先の設定などを行うこともできる。

【0033】

管理端末160は記憶デバイス制御装置100に内蔵されている形態とすることもできるし、外付けされている形態とすることもできる。また管理端末160は、記憶デバイス制御装置100及び記憶デバイス300の保守・管理を専用に行うコンピュータとすることもできるし、汎用のコンピュータに保守・管理機能を持たせたものとすることもできる。

【0034】

管理端末160の構成を示すブロック図を図2に示す。

管理端末160は、CPU161、メモリ162、ポート163、記録媒体読取装置164、入力装置165、出力装置166、記憶装置168を備える。

【0035】

CPU161は管理端末160の全体の制御を司るもので、メモリ162に格納されたプログラム162cを実行することにより上記Webサーバとしての機能等を実現する。メモリ162には、物理ディスク管理テーブル162aとLU管理テーブル162bとプログラム162cとが記憶されている。

【0036】

物理ディスク管理テーブル162aは、記憶デバイス300に備えられる物理ディスク（ディスクドライブ）を管理するためのテーブルである。物理ディスク管理テーブル162aを図3に示す。図3においては、記憶デバイス300が備える多数の物理ディスクのうち、ディスク番号#001乃至#006までが示されている。それぞれの物理ディスクに対して、容量、RAID構成、使用状況が

示されている。

LU管理テーブル162bは、上記物理ディスク上に論理的に設定されるLUを管理するためのテーブルである。LU管理テーブル162bを図4に示す。図4においては、記憶デバイス300上に設定される多数のLUのうち、LU番号#1乃至#3までが示されている。それぞれのLUに対して、物理ディスク番号、容量、RAID構成が示されている。

【0037】

記録媒体読取装置164は、記録媒体167に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ162や記憶装置168に格納される。従って、例えば記録媒体167に記録されたプログラム162cを、記録媒体読取装置164を用いて上記記録媒体167から読み取って、メモリ162や記憶装置168に格納するようにすることができる。記録媒体167としてはフレキシブルディスクやCD-ROM、DVD-ROM、DVD-RAM、半導体メモリ等を用いることができる。なお、上記プログラム162cは管理端末160を動作させるためのプログラムとすることができる他、チャンネル制御部110やディスク制御部140にOS701やアプリケーションプログラムをインストールするためのプログラムや、バージョンアップするためのプログラムとすることもできる。

【0038】

記録媒体読取装置164は管理端末160に内蔵されている形態とすることもできるし、外付されている形態とすることもできる。記憶装置168は、例えばハードディスク装置やフレキシブルディスク装置、半導体記憶装置等である。入力装置165はオペレータ等による管理端末160へのデータ入力等のために用いられる。入力装置165としては例えばキーボードやマウス等が用いられる。出力装置166は情報を外部に出力するための装置である。出力装置166としては例えばディスプレイやプリンタ等が用いられる。ポート163は内部LAN151に接続されており、これにより管理端末160はチャンネル制御部110やディスク制御部140等と通信を行うことができる。またポート163は、LA

N400に接続するようにすることもできるし、電話回線に接続するようにすることもできる。

【0039】

===外観図===

次に、本実施の形態に係るストレージシステム600の外観構成を図5に示す。また、記憶デバイス制御装置100の外観構成を図6に示す。

図5に示すように、本実施の形態に係るストレージシステム600は記憶デバイス制御装置100及び記憶デバイス300がそれぞれの筐体に納められた形態をしている。記憶デバイス制御装置100の筐体の両側に記憶デバイス300の筐体が配置されている。

【0040】

記憶デバイス制御装置100は、正面中央部に管理端末160が備えられている。管理端末160はカバーで覆われており、図6に示すようにカバーを開けることにより管理端末160を使用することができる。なお図6に示した管理端末160はいわゆるノート型パーソナルコンピュータの形態をしているが、どのような形態とすることも可能である。

【0041】

管理端末160の下部には、チャンネル制御部110のボードを装着するためのスロットが設けられている。チャンネル制御部110のボードとは、チャンネル制御部110の回路基板が形成されたユニットであり、スロットへの装着単位である。本実施の形態に係るストレージシステム600においては、スロットは8つあり、図5及び図6には8つのスロットにチャンネル制御部110のボードが装着された状態が示されている。各スロットにはチャンネル制御部110のボードを装着するためのガイドレールが設けられている。ガイドレールに沿ってチャンネル制御部110のボードをスロットに挿入することにより、チャンネル制御部110のボードを記憶デバイス制御装置100に装着することができる。また各スロットに装着されたチャンネル制御部110のボードは、ガイドレールに沿って手前方向に引き抜くことにより取り外すことができる。また各スロットの奥手方向正面部には、各チャンネル制御部110のボードを記憶デバイス制御装置100と電氣的に

接続するためのコネクタが設けられている。チャンネル制御部 110 には、CHN、CHF、CHA があるが、いずれのチャンネル制御部 110 のボードもサイズやコネクタの位置、コネクタのピン配列等に互換性をもたせているため、8つのスロットにはいずれのチャンネル制御部 110 のボードも装着することが可能である。従って、例えば8つのスロット全てにCHN 110のボードを装着するようにすることもできる。また図1に示したように、4枚のCHN 110のボードと、2枚のCHF 110のボードと、2枚のCHA 110のボードとを装着するようにすることもできる。チャンネル制御部 110 のボードを装着しないスロットを設けるようにすることもできる。

【0042】

各スロットのチャンネル制御部 110 は、同種の複数のチャンネル制御部 110 でクラスタを構成する。例えば2枚のCHN 110をペアとしてクラスタを構成することができる。クラスタを構成することにより、クラスタ内のあるチャンネル制御部 110 に障害が発生した場合でも、障害が発生したチャンネル制御部 110 がそれまで行っていた処理をクラスタ内の他のチャンネル制御部 110 に引き継ぐようにすることができる（フェイルオーバー制御）。2枚のCHN 110でクラスタを構成している様子を示す図を図12に示すが、詳細は後述する。

【0043】

なお、記憶デバイス制御装置 100 は信頼性向上のため電源供給が2系統化されており、チャンネル制御部 110 のボードが装着される上記8つのスロットは電源系統毎に4つずつに分けられている。そこでクラスタを構成する場合には、両方の電源系統のチャンネル制御部 110 のボードを含むようにする。これにより、片方の電源系統に障害が発生し電力の供給が停止しても、同一クラスタを構成する他方の電源系統に属するチャンネル制御部 110 のボードへの電源供給は継続されるため、当該チャンネル制御部 110 に処理を引き継ぐ（フェイルオーバー）ことができる。

【0044】

なお、上述したように、チャンネル制御部 110 は上記各スロットに装着可能なボードとして提供されるが、上記一つのボードは一体形成された複数枚数の回路

基板から構成されているようにすることもできる。

【0045】

ディスク制御部140や共有メモリ120等の、記憶デバイス制御装置100を構成する他の装置については図5及び図6には示されていないが、記憶デバイス制御装置100の背面側等に装着されている。

【0046】

また記憶デバイス制御装置100には、チャネル制御部110のボード等から発生する熱を放出するためのファン170が設けられている。ファン170は記憶デバイス制御装置100の上面部に設けられる他、チャネル制御部110用スロットの上部にも設けられている。

【0047】

ところで、筐体に収容されて構成される記憶デバイス制御装置100および記憶デバイス300としては、例えばSAN対応として製品化されている従来構成の装置を利用することができる。特に上記のようにCHN110のボードのコネクタ形状を従来構成の筐体に設けられているスロットにそのまま装着できる形状とすることで従来構成の装置をより簡単に利用することができる。つまり本実施例のストレージシステム600は、既存の製品を利用することで容易に構築することができる。

【0048】

===チャネル制御部===

本実施の形態に係るストレージシステム600は、上述の通りCHN110により情報処理装置1乃至3(200)からのファイルアクセス要求を受け付け、NASとしてのサービスを情報処理装置1乃至3(200)に提供する。

CHN110のハードウェア構成を図7に示す。この図に示すようにCHN110のハードウェアは一体的にユニット化されたボードで構成される。以下、このボードのことをNASボードとも記す。NASボードは一枚もしくは複数枚の回路基板を含んで構成される。より具体的には、NASボードは、ネットワークインタフェース部111、CPU112、メモリ113、入出力制御部114、I/O(Input/Output)プロセッサ119、NVRAM(Non Volatile RAM)1

15、ボード接続用コネクタ116、通信コネクタ117を備え、これらが同一のボードに形成されて構成されている。

【0049】

ネットワークインタフェース部111は、情報処理装置200との間で通信を行うための通信インタフェースを備えている。CHN110の場合は、例えばTCP/IPプロトコルに従って情報処理装置200から送信されたファイルアクセス要求を受信する。通信コネクタ117は情報処理装置200と通信を行うためのコネクタである。CHN110の場合はLAN400に接続可能なコネクタであり、例えばイーサネット（登録商標）に対応している。

CPU112は、CHN110をNASボードとして機能させるための制御を司る。

メモリ113には様々なプログラムやデータが記憶される。例えば図9に示すメタデータ730やロックテーブル720、また図11に示すNASマネージャ706等の各種プログラムが記憶される。

【0050】

メタデータ730はファイルシステムプログラム703等により実現されるファイルシステムが管理しているファイルに対応させて生成される情報である。メタデータ730には、例えばファイルのデータが記憶されているLU上のアドレスやデータサイズなど、ファイルの保管場所を特定するための情報が含まれる。メタデータ730にはファイルシステムが管理しているファイルのファイル名とそのファイルに対するデータが記憶されているLU上の位置を特定するためのアドレス情報（例えば先頭アドレス）とを対応させた情報が含まれている。さらにメタデータ730には、ファイルの容量、所有者、更新時刻等の情報が含まれることもある。また、メタデータ730はファイルだけでなくディレクトリに対応させて生成されることもある。メタデータ730の例を図13に示す。メタデータ730は記憶デバイス300上の各LUにも記憶されている。

【0051】

例えば情報処理装置200からあるCHN110に対するファイルアクセス要求があった場合、そのCHN110はそのファイルアクセス要求の対象として設

定されているファイルのファイル名をメタデータ730に対照することによりそのファイルが記憶されているLU上のアドレスを特定する。図14はCHN110がファイルアクセス要求に設定されているファイル名に対応するLU上のアドレスを特定する様子を説明している。この特定に際しては、CHN110のメモリ113上に記憶されているメタデータ730が参照される。なお、CHN110のメモリ113に記憶されているデータはメモリ113の効率的な利用のために、FIFO (First In First Out) 等に従ってメモリ113からキャッシュメモリ130に書き出されるように制御されることもある。ここでこの制御においてはキャッシュメモリ130には複数のCHN110から書き出されたメタデータ730が共存することもありうる。そのため、キャッシュメモリ130に記憶されているメタデータ730には、それぞれがいずれのCHN110から書き出されたものであるのかを特定するための識別子が付与されている。そして、CHN110は、キャッシュメモリ130に記憶されているメタデータ730を自身のメモリ113上に書き戻すに際しては、上記識別子を参照して自身の識別子が付与されている確認した場合にそのメタデータ730を書き戻す。

【0052】

ロックテーブル720は、情報処理装置1乃至3(200)からのファイルアクセスに対して排他制御を行うためのテーブルである。排他制御を行うことにより情報処理装置1乃至3(200)でファイルを共用することができる。ロックテーブル720を図15に示す。この図に示すように、ロックテーブル720にはファイルロックテーブル721とLUロックテーブル722とがある。ファイルロックテーブル721は、ファイル毎にロックが掛けられているか否かを示すためのテーブルである。いずれかの情報処理装置200によりあるファイルがオープンされている場合に当該ファイルにロックが掛けられる。ロックが掛けられたファイルに対する他の情報処理装置200によるアクセスは禁止される。LUロックテーブル722は、LU毎にロックが掛けられているか否かを示すためのテーブルである。いずれかの情報処理装置200により、あるLUに対するアクセスが行われている場合に当該LUにロックが掛けられる。ロックが掛けられたLUに対する他の情報処理装置200によるアクセスは禁止される。

【0053】

入出力制御部114は、ディスク制御部140やキャッシュメモリ130、共有メモリ120、管理端末160との間でデータやコマンドの授受を行う。入出力制御部114はI/Oプロセッサ119やNVRAM115を備えている。I/Oプロセッサ119は例えば1チップのマイコンで構成される。I/Oプロセッサ119は上記データやコマンドの授受を制御し、CPU112とディスク制御部140との間の通信を中継する。NVRAM115はI/Oプロセッサ119の制御を司るプログラムを格納する不揮発性メモリである。NVRAM115に記憶されるプログラムの内容は、管理端末160や、後述するNASマネージャ706からの指示により書き込みや書き換えを行うことができる。

【0054】

次にCHF110及びCHA110のハードウェア構成を示す図を図8に示す。CHF110やCHA110のボードも、CHN110と同様に一体的にユニット化されたボードとして形成されている。CHN110のボードと同様、上記ボードは一体形成された複数枚数の回路基板から構成されているようにすることもできる。またCHF110のボード及びCHA110のボードは、CHN110のボードとサイズやボード接続用コネクタ116の位置、ボード接続用コネクタ116のピン配列等に互換性をもたせている。

CHF110及びCHA110は、ネットワークインタフェース部111、メモリ113、入出力制御部114、I/Oプロセッサ119、NVRAM115、ボード接続用コネクタ116、通信コネクタ117を備える。

【0055】

ネットワークインタフェース部111は、情報処理装置200との間で通信を行うための通信インタフェースを備えている。CHF110の場合は、例えばファイバチャネルプロトコルに従って情報処理装置200から送信されたブロックアクセス要求を受信する。CHA110の場合は、例えばFICON（登録商標）やESCON（登録商標）、ACONARC（登録商標）、FIBARC（登録商標）のプロトコルに従って情報処理装置200から送信されたブロックアクセス要求を受信する。通信コネクタ117は情報処理装置200と通信を行うた

めのコネクタである。CHF110の場合はSAN500に接続可能なコネクタであり、例えばファイバチャネルに対応している。CHA110の場合は情報処理装置5と接続可能なコネクタであり、FICON（登録商標）やESCON（登録商標）、ACONARC（登録商標）、FIBARC（登録商標）に対応している。

【0056】

入出力制御部114は、それぞれCHF110、CHA110の全体の制御を司ると共に、ディスク制御部140やキャッシュメモリ130、共有メモリ120、管理端末160との間でデータやコマンドの授受を行う。メモリ113に格納された各種プログラムを実行することにより本実施の形態に係るCHF110及びCHA110の機能が実現される。入出力制御部114はI/Oプロセッサ119やNVRAM115を備えている。I/Oプロセッサ119は上記データやコマンドの授受を制御する。NVRAM115はI/Oプロセッサ119の制御を司るプログラムを格納する不揮発性メモリである。NVRAM115に記憶されるプログラムの内容は、管理端末160や、後述するNASマネージャ706からの指示により書き込みや書き換えを行うことができる。

【0057】

次にディスク制御部140のハードウェア構成を示す図を図10に示す。

ディスク制御部140は、一体的にユニット化されたボードとして形成されている。ディスク制御部140のボードは、インタフェース部141、メモリ143、CPU142、NVRAM144、ボード接続用コネクタ145を備え、これらが一体的にユニット化された回路基板として形成されている。

インタフェース部141は、接続部150を介してチャネル制御部110等との間で通信を行うための通信インタフェースや、記憶デバイス300との間で通信を行うための通信インタフェースを備えている。

【0058】

CPU142は、ディスク制御部140全体の制御を司ると共に、チャネル制御部140や記憶デバイス300、管理端末160との間の通信を行う。メモリ143やNVRAM144に格納された各種プログラムを実行することにより本

実施の形態に係るディスク制御部140の機能が実現される。ディスク制御部140により実現される機能としては、記憶デバイス300の制御やRAID制御、記憶デバイス300に記憶されたデータの複製管理やバックアップ制御、リモートコピー制御等である。

【0059】

NVRAM144はCPU142の制御を司るプログラムを格納する不揮発性メモリである。NVRAM144に記憶されるプログラムの内容は、管理端末160や、NASマネージャ706からの指示により書き込みや書き換えを行うことができる。

【0060】

またディスク制御部140のボードはボード接続用コネクタ145を備えている。ボード接続用コネクタ145が記憶デバイス制御装置100側のコネクタと嵌合することにより、ディスク制御部140のボードは記憶デバイス制御装置100と電氣的に接続される。

【0061】

===ソフトウェア構成===

次に、本実施の形態に係るストレージシステム600におけるソフトウェア構成図を図11に示す。

オペレーティングシステム701上では、RAIDマネージャ708、ボリュームマネージャ707、SVPマネージャ709、ファイルシステムプログラム703、ネットワーク制御部702、バックアップ管理プログラム710、障害管理プログラム705、NASマネージャ706などのソフトウェアが動作する。

【0062】

オペレーティングシステム701上で動作するRAIDマネージャ708は、RAID制御部740に対するパラメータの設定やRAID制御部740を制御する機能を提供する。RAIDマネージャ708はオペレーティングシステム701やオペレーティングシステム701上で動作する他のアプリケーション、もしくはSVPからパラメータや制御指示情報を受け付けて、受け付けたパラメー

タのRAID制御部740への設定や、RAID制御部指示情報に対応する制御コマンドの送信を行う。

【0063】

ここで設定されるパラメータとしては、例えば、RAIDグループを構成する記憶デバイス（物理ディスク）を定義（RAIDグループの構成情報、ストライプサイズの指定など）するためのパラメータ、RAIDレベル（例えば0, 1, 5）を設定するためのパラメータなどがある。また、RAIDマネージャ708がRAID制御部740に送信する制御コマンドとしてはRAIDの構成・削除・容量変更を指示するコマンド、各RAIDグループの構成情報を要求するコマンドなどがある。

【0064】

ボリュームマネージャ707は、RAID制御部740によって提供されるLUをさらに仮想化した仮想化論理ボリュームをファイルシステムプログラム703に提供する。1つの仮想化論理ボリュームは1以上の論理ボリュームによって構成される。

【0065】

ファイルシステムプログラム703の主な機能は、ネットワーク制御部702が受信したファイルアクセス要求に指定されているファイル名とそのファイル名が格納されている仮想化論理ボリューム上のアドレスとの対応づけを管理することである。例えば、ファイルシステムプログラム703はファイルアクセス要求に指定されているファイル名に対応する仮想化論理ボリューム上のアドレスを特定する。

【0066】

ネットワーク制御部702は、NFS（Network File System）711とSamba 712の2つのファイルシステムプロトコルを含んで構成される。NFS 711は、NFS 711が動作するUNIX（登録商標）系の情報処理装置200からのファイルアクセス要求を受け付ける。一方、Samba 712はCIFS（Common Interface File System）713が動作するWindows（登録商標）系の情報処理装置200からのファイルアクセス要求を受け付ける。

【0067】

NAS マネージャ 706 は、ストレージシステム 600 について、その動作状態の確認、設定や制御などを行うためのプログラムである。NAS マネージャ 706 は Web サーバとしての機能も有し、情報処理装置 200 からストレージシステム 600 の設定や制御を行うための設定 Web ページを情報処理装置 200 に提供する。設定 Web ページはチャンネル制御部 1 乃至 4 (110) の個々において動作する NAS マネージャ 706 により提供される。NAS マネージャ 706 は、情報処理装置 1 乃至 3 (200) からの HTTP (HyperText Transport Protocol) リクエストに応じて、設定 Web ページのデータを情報処理装置 1 乃至 3 (200) に送信する。情報処理装置 1 乃至 3 (200) に表示された設定 Web ページを利用してシステムアドミニストレータなどによりストレージシステム 600 の設定や制御の指示が行われる。

【0068】

NAS マネージャ 706 は、設定 Web ページに対する操作に起因して情報処理装置 200 から送信される設定や制御に関するデータを受信してそのデータに対応する設定や制御を実行する。これにより、情報処理装置 1 乃至 3 (200) からストレージシステム 600 の様々な設定や制御を行うことができる。また NAS マネージャ 706 は設定 Web ページの設定内容をチャンネル制御部 110 上で動作する OS やアプリケーションプログラム、ディスク制御部 14 等に通知する。設定 Web ページで設定された内容は共有 LU 310 に管理されることもある。

【0069】

NAS マネージャ 706 の設定 Web ページを利用して行うことができる内容としては、例えば、LU の管理や設定 (容量管理や容量拡張・縮小、ユーザ割り当て等)、上述の複製管理やリモートコピー (レプリケーション) 等の機能に関する設定や制御 (複製元の LU と複製先の LU の設定など)、後述のバックアップ管理プログラム 710 についての設定や制御、冗長構成された CHN や CHF、CHA 間でのクラスタの管理 (フェイルオーバーさせる相手の対応関係の設定、フェイルオーバー方法など)、OS や OS 上で動作するアプリケーションプログラ

ムのバージョン管理、ウイルス検知プログラムやウイルス駆除などのデータの安全性に関する機能を提供するセキュリティ管理プログラム716の動作状態の管理や設定などがある。

【0070】

バックアップ管理プログラム710は、記憶デバイス300に記憶されているデータをLAN経由またはSAN経由でバックアップするためのプログラムである。バックアップ管理プログラム710はNDMP (Network Data Management Protocol) の機能を提供し、情報処理装置200で動作するNDMP対応のバックアップソフトウェアとLAN400を通じてNDMPに従った通信を行う。バックアップデバイス910が情報処理装置200にSCSI経由などで接続されている場合、バックアップされるデータは情報処理装置200に一旦取り込まれてからバックアップデバイス910に送られる。バックアップデバイス910がLAN400接続されている場合には、バックアップされるデータを、情報処理装置200を経由せずにストレージシステム600から直接バックアップデバイス910に転送することもできる。

【0071】

障害管理プログラム705は、クラスタを構成するチャンネル制御部110間でのフェイルオーバー制御を行うためのプログラムである。

2枚のCHN110でクラスタ180が構成されている様子を示す図を図12に示す。図12では、CHN1 (チャンネル制御部1) 110とCHN2 (チャンネル制御部2) 110とでクラスタ180が構成されている場合を示す。

【0072】

上述したように、フェイルオーバー処理はクラスタ180を構成するチャンネル制御部110間で行われる。つまり、例えばCHN1 (110) に何らかの障害が発生し処理を継続することができなくなった場合には、CHN1 (110) がそれまで行っていた処理はCHN2 (110) に引き継がれる。フェイルオーバー処理は、CHN1 (110) とCHN2 (110) により実行される障害管理プログラム705により実行される。

【0073】

CHN1(110)及びCHN2(110)は共に障害管理プログラムを実行し、例えば共有メモリ120に対して自己の処理が正常に行われていることを書き込むようにする。そして、相手側の上記書き込みの有無を相互に確認するようにする。相手側による書き込みが検出できない場合には、相手側に何らかの障害が発生したと判断し、フェイルオーバ処理を実行する。フェイルオーバ実行時の処理の引き継ぎは、例えば共有LU310を介して行われる。

【0074】

フェイルオーバはこのように自動的に行われることもあるが、オペレータが管理端末を操作して手動で行われることもある。またユーザがNASマネージャ706が提供する設定Webページを利用して情報処理装置200から手動で行われることもある。フェイルオーバを手動で行う目的としては、耐用年数の経過やバージョンアップ、定期診断などのためにチャネル制御部110のハードウェア(例えばNASボード)を交換する必要がある場合などがある。

【0075】

SVPマネージャ709は、管理端末160からの要求に応じて各種のサービスを管理端末160に提供する。例えば、LUの設定内容やRAIDの設定内容等のストレージシステム600に関する各種設定内容の管理端末160への提供や、管理端末160から入力されたストレージシステム600に関する各種設定の反映等を行う。

【0076】

セキュリティ管理プログラム716は、コンピュータウイルスの検知、侵入監視、コンピュータウイルス検知プログラムの更新管理、感染したコンピュータウイルスの駆除、ファイアウォール機能などを実現する。

【0077】

===複製管理機能===

複製管理機能はあるLU(以下、「複製元LU」と記す)に記憶されているデータの複製をこれとは別のLU(以下、「複製先LU」と記す)にも記憶する機能である。複製管理機能は、ディスク制御部140のCPU142がNVRAM144に記憶されている複製管理プログラム760を実行することにより実現さ

れる。

【0078】

図16は複製管理機能に関する処理を説明するフローチャートである。複製元LUと複製先LUとの対応づけは、NASマネージャ706が提供する設定Webページを利用することにより情報処理装置200から設定することができる。図17に複製元のデータが記憶される複製元LUと複製先のデータが記憶される複製先LUとの対応づけ（ペア）を設定する際に利用される設定Webページの一例を示している。対応づけの設定に際しては、この設定Webページ1700の複製元LUの欄に複製元LUとするLUNを設定しその右側の複製先LUの欄に複製先LUとするLUNを設定した後（S1611）、「OK」ボタンをクリックする。これにより設定Webページに設定された内容が情報処理装置200からNASマネージャ706に送信される（S1612）。NASマネージャ706は設定内容を受信すると（S1613）、その設定内容をディスク制御部140に送信する（S1614）。ディスク制御部140は設定内容を受信するとこれをNVRAM144に記憶する（S1615）。複製管理プログラム760はNVRAM144に記憶されている複製元LUと複製先LUとの関係に従って複製管理機能を実行している。以上の処理の後、新たに設定された複製元LUと複製先LUとの間での複製管理機能が開始される（S1616）。

【0079】

ところで、管理端末160やNASマネージャ706からの指示に応じて複製元LUと複製先LUの制御状態は「ペア状態」（複製管理処理の実行状態）および「スプリット状態」（複製管理処理の停止状態）の双方向に移行させることができる。「ペア状態」にある複製元LUと複製先LUの関係においては、複製元LUの内容が更新されると、複製先LUの内容も直ちに更新される。すなわち、「ペア状態」においては複製元LUと複製先LUの間では、双方の内容の同一性がリアルタイムに確保されることになる。「スプリット状態」にある複製元LUと複製先LUの関係においては、複製元LUに対して更新があった場合でも、その更新内容は複製先LUに直ぐには反映されず、「スプリット状態」にある間に複製元LUに書き込まれたデータは再び「ペア状態」へと移行する際に複製先

L Uに反映される。なお、「スプリット状態」の間における複製元L Uの内容と複製先L Uの内容との差分のデータは複製管理機能により管理される。この差分データは例えばL Uの記憶領域上に設定される領域管理単位であるブロックやトラックごとに管理される。

【 0 0 8 0 】

図 1 8 は「ペア状態」から「スプリット状態」への移行を指示する際に利用される設定Web ページの一例である。この設定Web ページ 1 8 0 0 では、上記の移行を複製元L Uと複製先L Uのペアごとに設定することができる。「ペア状態」から「スプリット状態」への移行に関する処理について図 1 8 に示すフローチャートとともに説明する。複製元L Uと複製先L Uとのペアについて「スプリット状態」への移行（複製管理処理の停止）を指示する場合には、図 2 1 の設定Web ページのスプリット欄に「実行」を設定する（S1911）。その後、「OK」ボタンがクリックされると、その旨が情報処理装置 2 0 0 からNAS マネージャ 7 0 6 に送信される（S1912）。そして、これを受信（S1913）したNAS マネージャ 7 0 6 は、ディスク制御部 1 4 0 に「スプリット状態」への移行を指示する設定がされている複製元L Uと複製先L Uとのペアについて「スプリット状態」への移行を指示するコマンド（以下、「スプリット指示」と記す）を送出する（S1914）。ディスク制御部 1 4 0 で動作する複製管理プログラム 7 6 0 は、前記コマンドを受信すると該当の複製元L Uと複製先L Uを「ペア状態」から「スプリット状態」に移行させる（S1915）。

【 0 0 8 1 】

なお、「ペア状態」から「スプリット状態」に移行することにより、複製先L Uの内容はスプリット状態に移行した時点の内容のまま保持される。「スプリット指示」は、例えば、データをバックアップする際に送出される。すなわち、「スプリット状態」にある複製先L Uのデータをバックアップすることにより、複製元L Uに影響を与えることなくバックアップを取得することができる。また「スプリット指示」は、過去のある時点における複製元L Uの内容にアクセスしたい場合にも送出される。すなわち、「スプリット状態」にある複製先L Uの内容は、「スプリット状態」に移行した時点の内容のまま保持されるので、ユーザは

過去のある時点における複製元LUの内容にアクセスすることができる。なお、このような用途のために送出されるスプリット指示は、「スナップショット指示」と称されることがある。また、複製先LUに保持される過去のある時点における複製元LUの内容は、スナップショットイメージと称されることがある。複数の時点におけるスナップショットイメージを随時にバックアップすることで複製元LUの内容を世代管理することができる。

【 0 0 8 2 】

バックアップが完了した場合や、スナップショットイメージを保持しておく必要が無くなった場合などには、「スプリット状態」から「ペア状態」への移行が行われる。この移行は、図 2 0 の設定 Web ページ 2 0 0 0 を利用して情報処理装置 2 0 0 から指示することができる。この場合の処理について図 2 1 に示すフローチャートとともに説明する。

【 0 0 8 3 】

複製元LUと複製先LUとのペアについて「ペア状態」への移行を指示（複製管理処理の再開）する場合には、図 2 0 の設定 Web ページ 2 0 0 0 のリシンク欄に「実行」を設定する（S2111）。そして「OK」ボタンがクリックされると、その旨が情報処理装置 2 0 0 からNASマネージャ 7 0 6 に送信される（S2112）。これを受信（S2113）したNASマネージャ 7 0 6 は、ディスク制御部 1 4 0 に該当の複製元LUと複製先LUについての「ペア状態」への移行を指示するコマンド（以下、「リシンク指示」と記す）を送出する（S2114）。ディスク制御部 1 4 0 が前記コマンドを受信すると、ディスク制御部 1 4 0 で動作する複製管理プログラム 7 6 0 は、管理していた差分データを利用して複製元LUと複製先LUの内容を一致させる（S2115）。そして複製元LUと複製先LUの内容が一致した後、複製管理プログラム 7 6 0 は、複製元LUと複製先LUとを「ペア状態」に移行させる（S2116）。

【 0 0 8 4 】

複製管理機能を適用するかどうかをファイル単位やディレクトリ単位で設定することもできる。複製管理機能の適用対象に設定されているファイルやディレクトリは、自動的に複製元LUに記憶もしくは作成されるようになる。図 2 2 はフ

ファイル単位やディレクトリ単位で複製管理機能を適用するかどうかを設定する場合に利用される設定Webページ2200の一例である。この画面では複製管理機能を適用しようとするファイル名やディレクトリ名を設定することができる。設定Webページに設定された内容は、「OK」ボタンがクリックされることによりNASマネージャ706に送信される。NASマネージャ706は受信した設定内容を共有メモリ120に記憶する。共有メモリ120の内容は、例えばストレージシステム600が情報処理装置200から複製管理機能が適用されるように設定されているファイルやディレクトリを対象とするファイルアクセス要求を受信した場合にファイルシステムプログラム703により参照される。この場合にファイルシステムプログラム703が行う処理について図23に示すフローチャートとともに説明する。

【0085】

ストレージシステム600がファイルアクセス要求を受信すると(S2311)、ファイルシステムプログラム703は共有メモリ120の内容を参照し、そのファイルアクセス要求で指定されるファイルやディレクトリが複製管理機能の適用対象となっているかどうかを調べる(S2312)。適用対象でなければ(S2312:NO)通常書き込み処理を実行する(S2313)。一方、複製管理機能の適用対象であった場合には(S2312:YES)、その書き込みデータの書き込み先が複製管理機能の複製元LUに設定されているLUの記憶領域となるようにメタデータ730を設定する(S2314)。これにより複製管理機能の適用対象のファイルは複製元LUに記憶されることになり、複製管理機能の適用対象のファイルの複製およびそのメタデータ730が自動的に複製先LUにも記憶されるようになる(S2315)。なお、複製管理機能により複製先LUにも自動的にメタデータ730が記憶される。従って複製先LUに記憶されているデータについてもファイルシステムプログラム703により管理することができる。

【0086】

ファイル単位やディレクトリ単位でスナップショットイメージを作成することができる。図2400はファイル単位やディレクトリ単位でスナップショットイメージを作成する際に利用される設定Webページの一例である。この設定Web

b ページ 2400 では、スナップショットを作成しようとするファイルやディレクトリおよびスナップショットを実行する日時を設定することができるようになっている。

【0087】

図 25 はファイル単位やディレクトリ単位でスナップショットイメージを作成する場合における処理を説明するフローチャートである。設定 Web ページにファイルやディレクトリが設定された後 (S2511)、設定 Web ページの「OK」ボタンがクリックされるとスナップショットの設定内容が NAS マネージャ 706 に送信される (S2512)。NAS マネージャ 706 は設定内容を受信すると (S2513)、設定内容として指定されているファイルやディレクトリが格納されている複製元 LU とその複製が管理されている複製先 LU との間のペアについての「スプリット指示」をディスク制御部 140 に送信する (S2514)。ディスク制御部 140 の複製管理プログラム 760 は「スプリット指示」を受信するとスプリットを実行し前記ペアを解除する (S2515)。これにより複製先 LU には設定内容で指定されるファイルやディレクトリについてのスナップショットイメージが保持されることになる。このように LU 単位だけでなくファイル単位やディレクトリ単位でスナップショットイメージを作成できることでより細かなサービスを提供することが可能となる。

【0088】

=== 整合性の確保 ===

ところで、データサイズが巨大である等の理由により、1つのファイルに対応するデータ（以下、「ファイルデータ」と記す）が複数の LU に跨って記憶されることもある。ここでこのようなファイルデータについての記憶デバイス 300 への書き込みが行われている途中において、その書き込み対象になっている複製元 LU 及びこれとペアを構成している複製先 LU との関係が「スプリット状態」に移行してしまうと、複製先 LU においてファイルデータ全体としての整合性が保証されなくなってしまう。そのため、本実施例のストレージシステム 600 は、そのような場合でもファイルデータ全体としての整合性が保証されるようにする仕組みを備えている。この仕組みについて図 26 に示すフローチャートとともに

に具体的に説明する。

【0089】

NAS マネージャ 706 はある複製元 LU と複製先 LU とのペアに対する「スプリット指示」を送出しようとする場合、OS が管理している記憶デバイス 300 へのデータ I/O 要求のキューに前記ペアの複製元 LU を対象とするデータ I/O 要求が含まれているかどうかを調べる (S2611)。ここで前記キューにそのようなデータ I/O 要求が含まれている場合には (S2611:YES)、さらに前記キューに他の複製元 LU に跨って記憶されている（もしくは、記憶される）ファイルを対象とするデータ I/O 要求が含まれているかどうかを調べる (S2612)。ここでそのようなデータ I/O 要求が含まれている場合には (S2612:YES)、NAS マネージャ 706 はそのデータ I/O 要求についての処理が完了するのを待ってから前記ペアに対する「スプリット指示」を複製管理プログラム 760 に送出する (S2613)。これにより複製先 LU においてもファイルデータ全体としての整合性が保証される。なお、(S2611:NO) または (S2612:NO) の場合には、上述した通常の仕組みによりスプリットに関する処理が行われる (S2614)。

【0090】

ところで、データ I/O 要求に対応して行われる記憶デバイス 300 への書き込み処理や読み出し処理は必ずしも全ての複製元 LU に対して同時に進行するわけではない。そこで、データ I/O 要求に対応する全ての処理が完了するまで待たずに、スプリットしようとするペアについての処理が完了するのを確認した時点で「スプリット指示」を送出するようにしてもよい。

【0091】

以上のように設定 Web ページを利用して情報処理装置 200 から複製管理機能についての設定や制御が行えることで、情報処理装置 200 を利用するユーザー側の視点から複製管理機能についての柔軟な運用が可能となる。また、LU 単位だけでなくファイル単位やディレクトリ単位で複製管理機能に関する設定や制御ができることで情報処理装置 200 のユーザーに対するサービスの向上が図られる。なお、以上に説明した設定や制御は、管理端末 160 から行うようにすることも可能である。

【0092】

===リモートコピー===

ディスク制御部140のCPU142がNVRAM144に記憶されているソフトウェアを実行することにより実現される上述のリモートコピー（遠隔複製又はレプリケーション）の機能について、NASマネージャ706により提供される設定Webページを利用してその設定や制御を行うこともできる。リモートコピーの機能が動作中は、ストレージシステム600における複製元のLU（以下、「プライマリLU」と記す）にデータが書き込まれると、そのデータがSAN500を介してストレージシステム600から記憶他方の装置システム610に送信され、そのデータがストレージシステム610のLU（以下、「セカンダリLU」と記す）にも書き込まれる。このようにリモートコピー機能の動作中は、プライマリLUとセカンダリLUの内容を一致させるように制御がなされる。

【0093】

リモートコピーの方式としては同期方式と非同期方式とがある。同期方式の場合、情報処理装置200からプライマリLUへのデータ書き込みを指示するデータ入出力要求を受信すると、ストレージシステム600はプライマリLUにそのデータ入出力要求に対応するデータを書き込む。またストレージシステム600は書き込んだデータと同じデータをストレージシステム610に送信する。ストレージシステム610はストレージシステム600から送信されてくるデータを受信すると、そのデータをセカンダリLUに書き込む。そしてストレージシステム610は、ストレージシステム610はデータを書き込んだ旨をストレージシステム600に対して通知する。そしてこの通知を受信したストレージシステム600は、情報処理装置200にデータの書き込みを完了した旨を通知するメッセージを送信する。

【0094】

このように同期方式の場合には、プライマリLUとセカンダリLUの双方にデータが書き込まれたことが確認された後にはじめて情報処理装置200に完了通知が送信される。このため、同期方式では情報処理装置200が完了通知を受信した時点において必ずプライマリLUの内容とセカンダリLUの内容の一致性が

確保されることになる。但し、同期方式の場合には、セカンダリLUへのデータの書き込みが完了するまでは情報処理装置200に完了通知が報告されない。従って、同期方式の場合にはストレージシステム600にアクセスする情報処理装置200からストレージシステム600にデータ入出力要求が送信されてから情報処理装置200に完了通知が返ってくるまでの間のレスポンスタイムが非同期方式の場合に比べると一般に長くなる。

【0095】

非同期方式の場合には、情報処理装置200からプライマリLUへのデータ書き込みを指示するデータ入出力要求を受信したストレージシステム600はこれに応じてプライマリLUにデータを書き込む。また、書き込んだデータと同じデータをストレージシステム610に送信する。ストレージシステム610はストレージシステム600から送信されてくるデータを受信すると、そのデータをセカンダリLUに書き込む。そしてデータを書き込んだ旨をストレージシステム600に通知する。ここでストレージシステム600は、プライマリLUにデータを書き込むと、ストレージシステム610にデータが書き込まれたかどうかとは関係なく情報処理装置200に対して上記データ入出力要求についての完了通知を送信してしまう。このため非同期方式では情報処理装置200へのレスポンスタイムが同期方式の場合に比べて一般に短くなる。但し非同期方式では同期方式のように情報処理装置200が完了通知を受信したとしてもその時点ではプライマリLUとセカンダリLUとの間のデータの一致性は必ずしも保証されない。なお、リモートコピーが適用される場合には、一般にプライマリLUとセカンダリLUとの内容の差分に関する情報がストレージシステム600において管理される。

【0096】

以上に説明したプライマリLUとセカンダリLUとの対応づけや、「同期方式」や「非同期方式」の設定などのリモートコピー機能に関する設定は、NASマネージャ706が提供する設定Webページを利用して情報処理装置200から行うことができる。リモートコピー機能の設定に関する処理について図27に示すフローチャートとともに説明する。この設定に際し利用される設定Webペー

ジの一例を図28に示している。

プライマリLUとセカンダリLUとを対応づける場合には、この設定Webページ2800のプライマリLUの欄にプライマリLUとするLUNを設定し、その右側のセカンダリLUの欄にセカンダリLUとするLUNを設定する。また、その右側にあるチェック欄をチェックすることにより、このリモートコピーペアを「同期方式」で運用するのか、「非同期方式」で運用するのかを設定することができる(S2711)。「OK」ボタンがクリックされると設定Webページに設定された内容が情報処理装置200からNASマネージャ706に送信される(S2712)。

【0097】

NASマネージャ706は設定内容を受信すると(S2713)これをディスク制御部140に送信する(S2714)。ディスク制御部140はこの設定内容を受信するとその内容に従ってメモリ143にプライマリLUとセカンダリLUとの対応づけおよびこのペアのリモートコピーの方式を記憶する(S2715)。

【0098】

リモートコピー制御プログラム750は、メモリ143に記憶されているプライマリLUとセカンダリLUとの関係やリモートコピーの方式に従ってリモートコピーを実行している。従って以上の処理の後には新たに設定されたプライマリLUの内容がセカンダリLUにも記憶されるように制御が開始される(S2716)。なおこの制御は設定されているリモートコピーの方式に従ってなされる。

【0099】

リモートコピー機能を適用するかどうかをファイル単位やディレクトリ単位で設定することもできる。この場合、リモートコピー機能が適用されるように設定されているファイルやディレクトリは自動的にプライマリLUに記憶もしくは作成されることになる。図29はリモートコピー機能を適用するかどうかをファイル単位やディレクトリ単位で設定する際に利用される設定Webページの一例である。この設定Webページ2900の「ファイル名」の欄にはリモートコピー機能を適用しようとするファイル名を設定する。この図の「ディレクトリ名」の欄にはリモートコピー機能を適用しようとするディレクトリ名を指定する。設定

Web ページ 2900 に設定された内容はこのページの「OK」ボタンがクリックされることにより NAS マネージャ 706 に送信される。NAS マネージャ 706 は設定内容を受信するとこれをメモリ 113 に記憶する。メモリ 113 に記憶された内容はストレージシステム 600 が情報処理装置 200 からリモートコピー機能が適用されるように設定されているファイルやディレクトリを対象とするファイルアクセス要求を受信した場合にファイルシステムプログラム 703 により参照される。

【0100】

次にこの場合にファイルシステムプログラム 703 が行う処理について、図 30 に示すフローチャートとともに説明する。

ストレージシステム 600 がファイルアクセス要求を受信すると (S3011)、ファイルシステムプログラム 703 はメモリ 143 の内容を参照し、そのファイルアクセス要求で指定されるファイルやディレクトリがリモートコピー機能の適用対象となっているかどうかを調べる (S3012)。ファイルやディレクトリがリモートコピー機能の適用対象であった場合には (S3012: YES)、その書き込みデータの書き込み先がリモートコピー機能のプライマリ LU に設定されている LU の記憶領域となるようにメタデータ 730 を設定する (S3013)。これによりリモートコピー機能の適用対象のファイルはプライマリ LU に記憶されることになり (S3014)、リモートコピー機能の適用対象のファイルの複製およびそのメタデータ 730 が自動的にセカンダリ LU にも記憶されるようになる。 (S3012: NO) の場合には、通常の手書き込み処理を実行する (S3015)。

【0101】

なお、リモートコピー機能によりセカンダリ LU にも自動的にメタデータ 730 が記憶される。従ってセカンダリ LU に記憶されているデータについてもファイルとしてファイルシステムが認識することができる。従ってストレージシステム 610 においてファイルシステムプログラムが動作している場合には、ストレージシステム 610 はセカンダリ LU に記憶されているデータをファイルとして認識することができる。またストレージシステム 610 がブロック単位でのアクセスのみしか受け付けない場合でも、例えばセカンダリ LU を提供しているスト

レージシステム 6 1 0 に直接アクセスしてくる情報処理装置 2 0 0 でファイルシステムプログラムが動作していれば、その情報処理装置 2 0 0 は上記メタデータ 7 3 0 を利用してセカンダリ L U に記憶されているデータをファイルとして認識することができる。

【 0 1 0 2 】

以上に説明したストレージシステム 6 0 0 においては、N A S マネージャ 7 0 6 が提供する設定 W e b ページを利用して情報処理装置 2 0 0 からリモートコピー機能についての設定や制御を行うことができる。これにより情報処理装置 2 0 0 を利用するユーザ側の視点からリモートコピー機能についての柔軟な運用が可能となり、サービスの向上が図られる。なお、以上に説明した設定や制御は、管理端末 1 6 0 から行うようにすることも可能である。

【 0 1 0 3 】

以上本実施の形態について説明したが、上記実施例は本発明の理解を容易にするためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に、本発明にはその等価物も含まれる。

【 0 1 0 4 】

【発明の効果】

記憶デバイス制御装置の制御方法、及び記憶デバイス制御装置を提供することができる。

【図面の簡単な説明】

【図 1】 本実施の形態に係るストレージシステムの全体構成を示すブロック図である。

【図 2】 本実施の形態に係る管理端末の構成を示すブロック図である。

【図 3】 本実施の形態に係る物理ディスク管理テーブルを示す図である。

【図 4】 本実施の形態に係る L U 管理テーブルを示す図である。

【図 5】 本実施の形態に係るストレージシステムの外観構成を示す図である。

【図 6】 本実施の形態に係る記憶デバイス制御装置の外観構成を示す図で

ある。

【図 7】 本実施の形態に係る C H N を示す図である。

【図 8】 本実施の形態に係る C H F、C H A を示す図である。

【図 9】 本実施の形態に係るメモリに記憶されるデータの内容を説明するための図である。

【図 1 0】 本実施の形態に係るディスク制御部を示す図である。

【図 1 1】 本実施の形態に係るソフトウェア構成図である。

【図 1 2】 本実施の形態に係るチャネル制御部においてクラスタが構成されている様子を示す図である。

【図 1 3】 本実施の形態に係るメタデータを示す図である。

【図 1 4】 本実施例に係る C H N がファイルアクセス要求に設定されているファイル名に対応する L U 上のアドレスを特定する様子を説明する図である。

【図 1 5】 本実施の形態に係るロックテーブルを示す図である。

【図 1 6】 本実施の形態に係る、複製管理機能に関する処理を説明するフローチャートである。

【図 1 7】 本実施の形態に係る、複製元 L U と複製先のデータが記憶される複製先 L U との対応づけを設定する際に利用される設定 W e b ページである。

【図 1 8】 本実施の形態に係る、「ペア状態」から「スプリット状態」への移行を指示する際に利用される設定 W e b ページである。

【図 1 9】 本実施の形態に係る、「ペア状態」から「スプリット状態」への移行に関する処理を説明するフローチャートである。

【図 2 0】 本実施の形態に係る、「スプリット状態」から「ペア状態」への移行を指示する際に利用される設定 W e b ページである。

【図 2 1】 本実施の形態に係る、「スプリット状態」から「ペア状態」への移行に関する処理を説明するフローチャートである。

【図 2 2】 本実施の形態に係る、ファイル単位やディレクトリ単位で複製管理機能を適用するかどうかを設定する場合に利用される設定 W e b ページである。

【図 2 3】 本実施の形態に係る、ファイル単位やディレクトリ単位で複製

管理機能を適用するかどうかを設定する場合における処理を説明するフローチャートである。

【図 2 4】 本実施の形態に係る、ファイル単位やディレクトリ単位でスナップショットイメージを作成する際に利用される設定 Web ページである。

【図 2 5】 本実施の形態に係る、ファイル単位やディレクトリ単位でスナップショットイメージを作成する場合における処理を説明するフローチャートである。

【図 2 6】 本実施の形態に係る、ファイルデータ全体としての整合性が保証されるようにする仕組みを説明するフローチャートである。

【図 2 7】 本実施の形態に係る、リモートコピー機能におけるプライマリ LU とセカンダリ LU との対応づけの設定に関する処理を説明するフローチャートである。

【図 2 8】 本実施の形態に係る、リモートコピー機能におけるプライマリ LU とセカンダリ LU との対応づけ等の設定に際し利用される設定 Web ページである。

【図 2 9】 本実施の形態に係る、リモートコピー機能を適用するかどうかをファイル単位やディレクトリ単位で設定する際に利用される設定 Web ページである。

【図 3 0】 本実施の形態に係る、リモートコピー機能が適用されるように設定されているファイルやディレクトリを対象とするファイルアクセス要求を受信した場合に行われる処理を説明するフローチャートである。

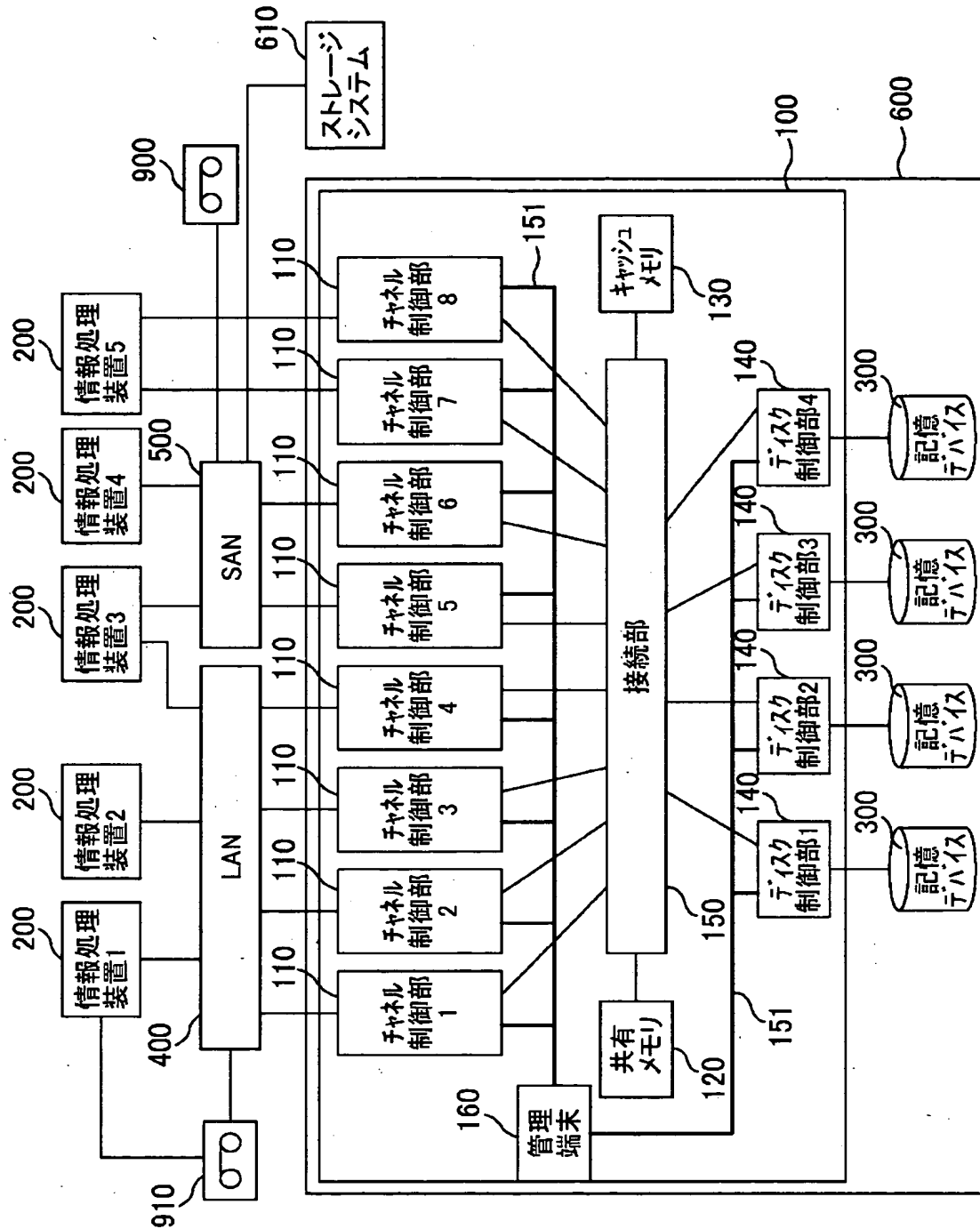
【符号の説明】

1 0 0 記憶デバイス制御装置	1 1 0 チャンネル制御部
1 1 1 ネットワークインタフェース部	1 1 2 CPU
1 1 3 メモリ	1 1 4 入出力制御部
1 1 5 NVRAM	1 1 6 ボード接続用コネクタ
1 1 7 通信コネクタ	1 1 8 回路基板
1 1 9 I/O プロセッサ	1 2 0 共有メモリ
1 3 0 キャッシュメモリ	1 4 0 ディスク制御部

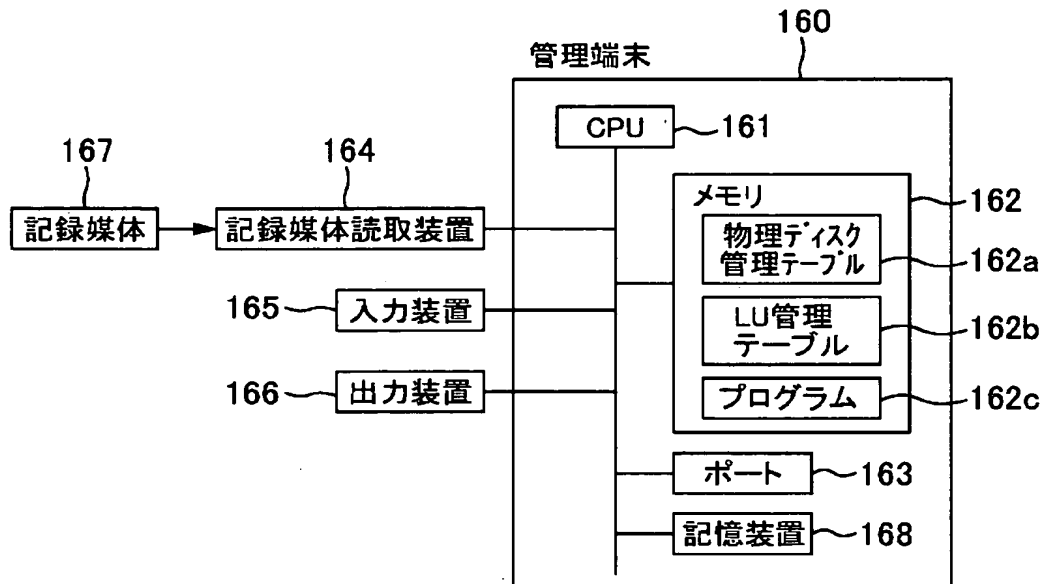
1 5 0	接続部	1 5 1	内部 L A N
1 6 0	管理端末	1 7 0	ファン
1 8 0	クラスタ	2 0 0	情報処理装置
3 0 0	記憶デバイス	4 0 0	L A N
5 0 0	S A N	6 0 0	ストレージシステム
7 3 0	メタデータ	7 2 1	ファイルロックテーブル
7 2 2	L U ロックテーブル	7 7 0	マイクロプログラム
7 7 1	ローダ	7 7 2	インストーラ
7 7 3	O S	9 0 0	S A N 対応テープデバイス
9 1 0	テープデバイス		

【書類名】 図面

【図 1】



【図 2】



【図 3】

162a 物理ディスク管理テーブル

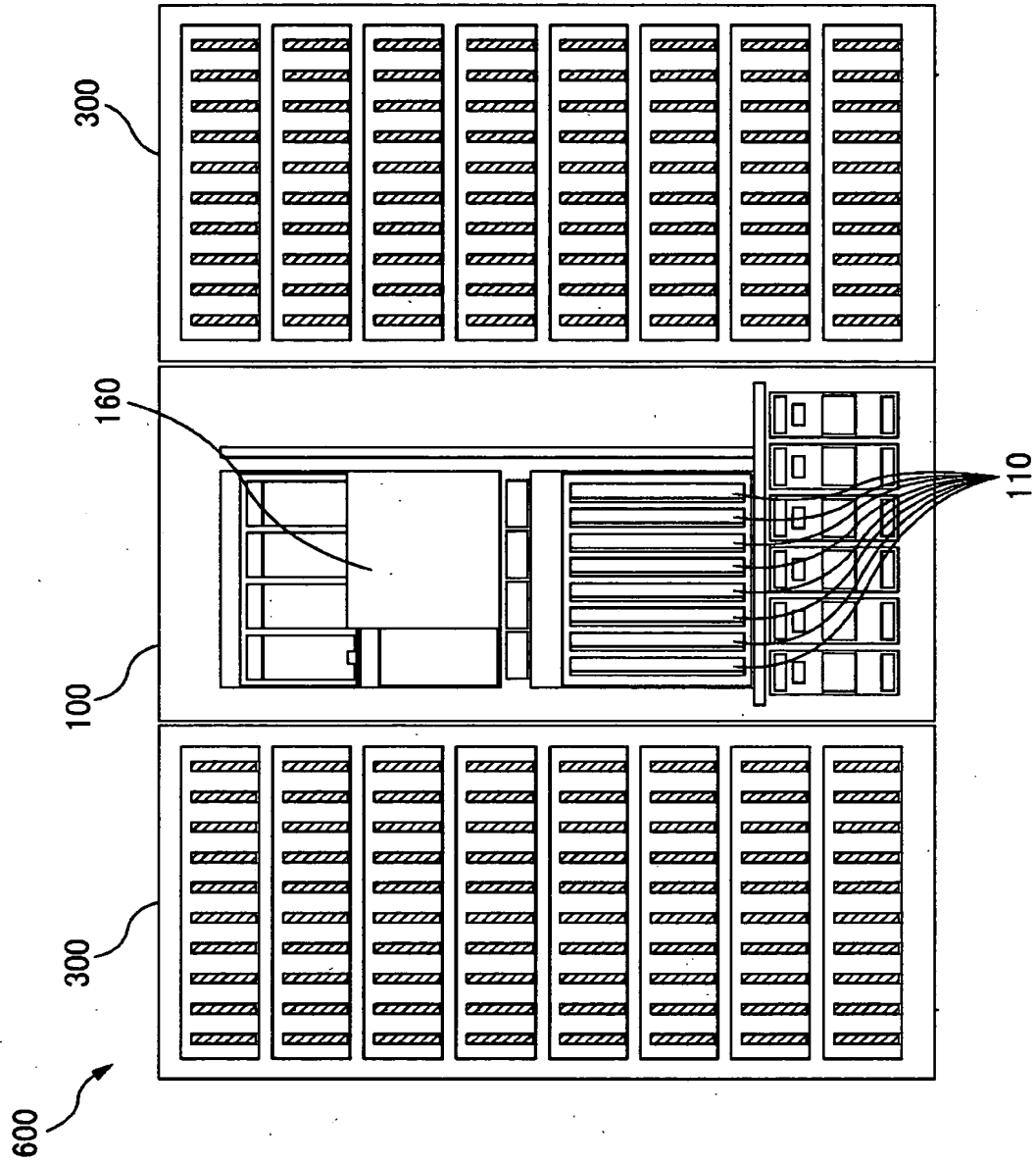
ディスク番号	容量	RAID	使用状況
#001	100GB	5	使用中
#002	100GB	5	使用中
#003	100GB	5	使用中
#004	100GB	5	使用中
#005	100GB	5	使用中
#006	50GB	—	未使用
⋮	⋮	⋮	⋮

【図 4】

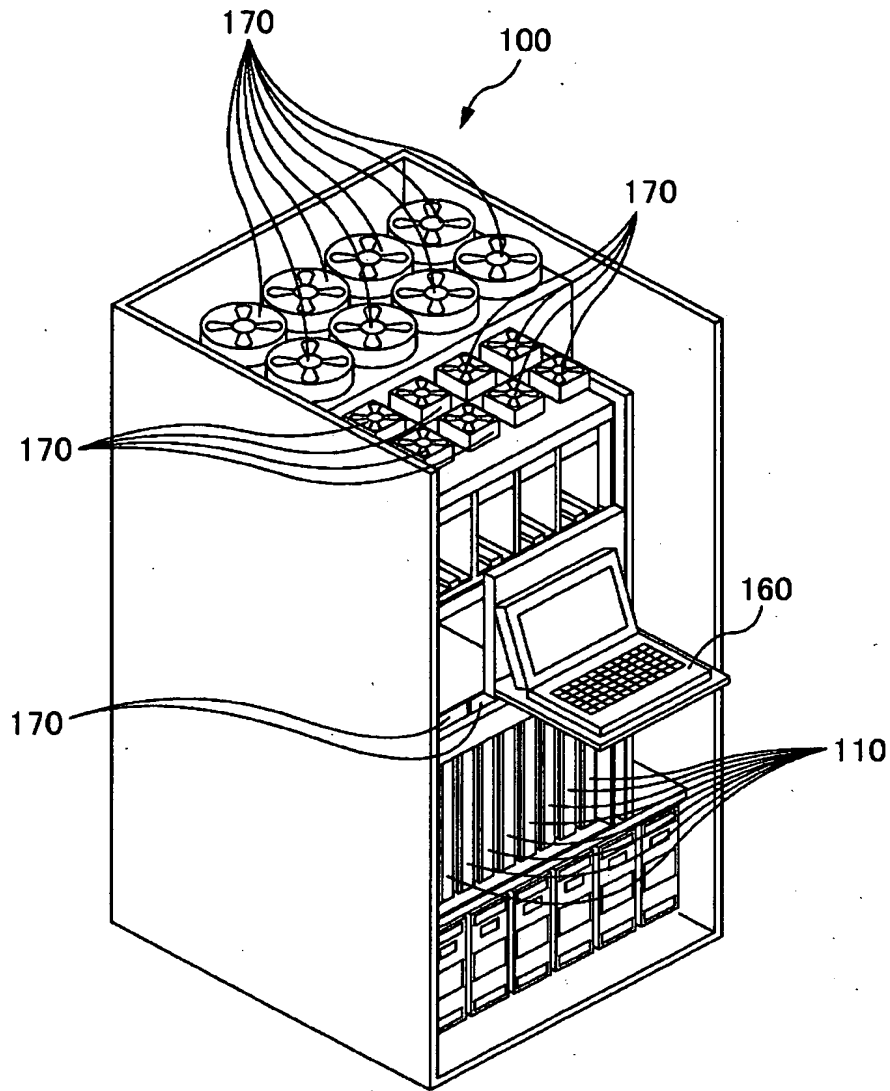
162b LU管理テーブル

LU番号	物理ディスク	容量	RAID
#1	#001,#002,#003,#004,#005	100GB	5
#2	#001,#002,#003,#004,#005	300GB	5
#3	#006,#007,	200GB	1
⋮	⋮	⋮	⋮

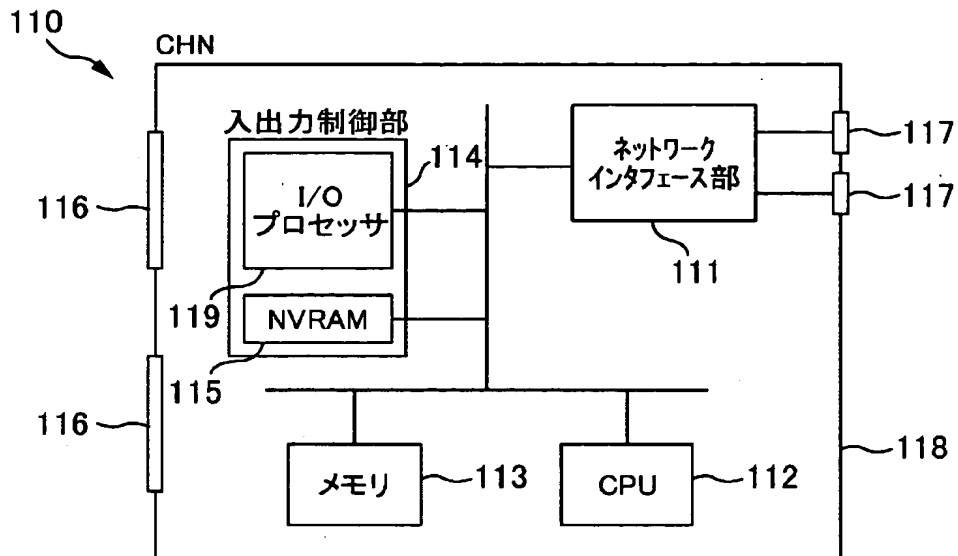
【图 5】



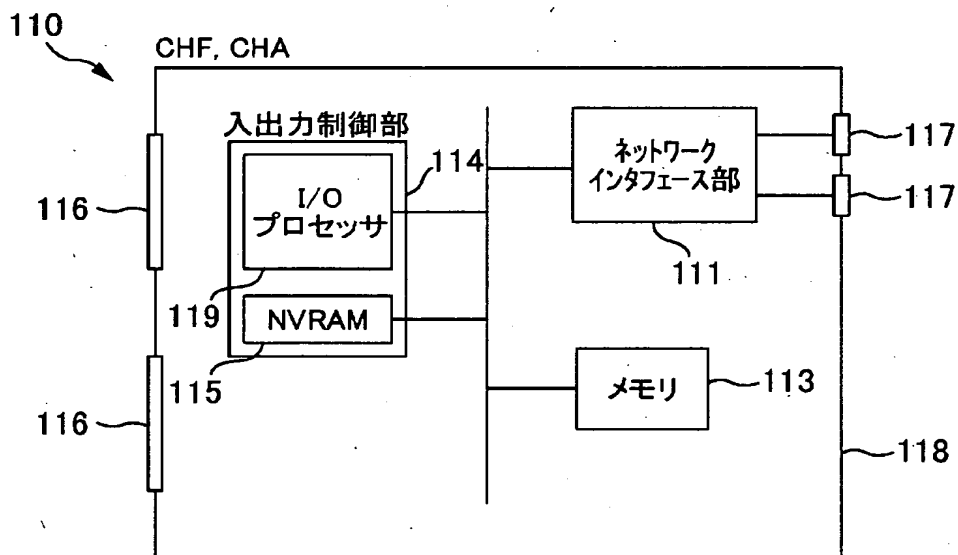
【図 6】



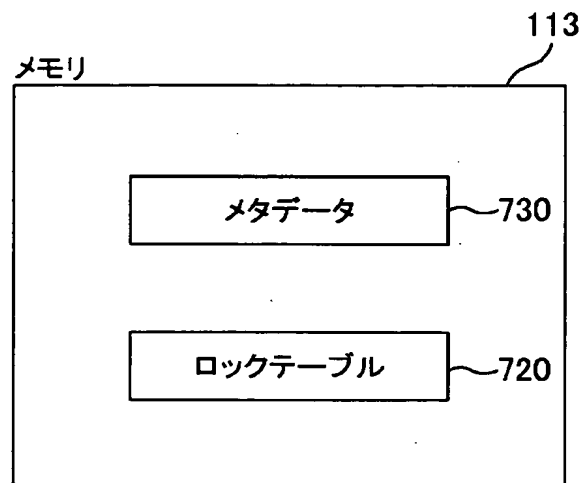
【図 7】



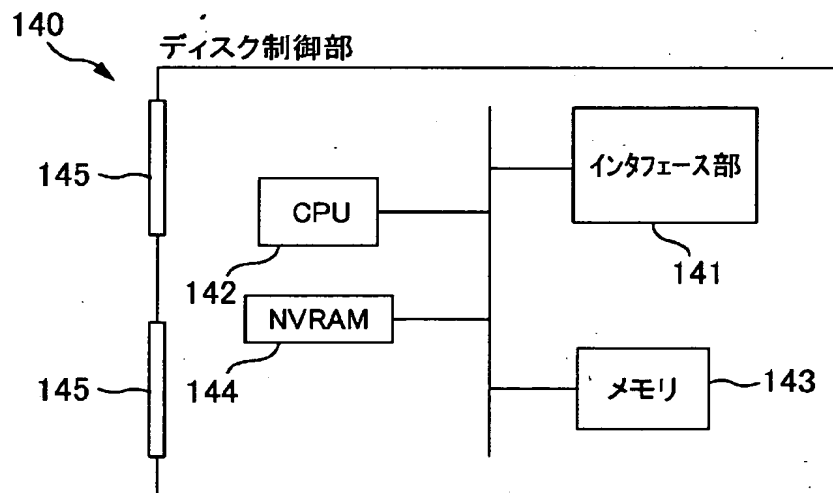
【図 8】



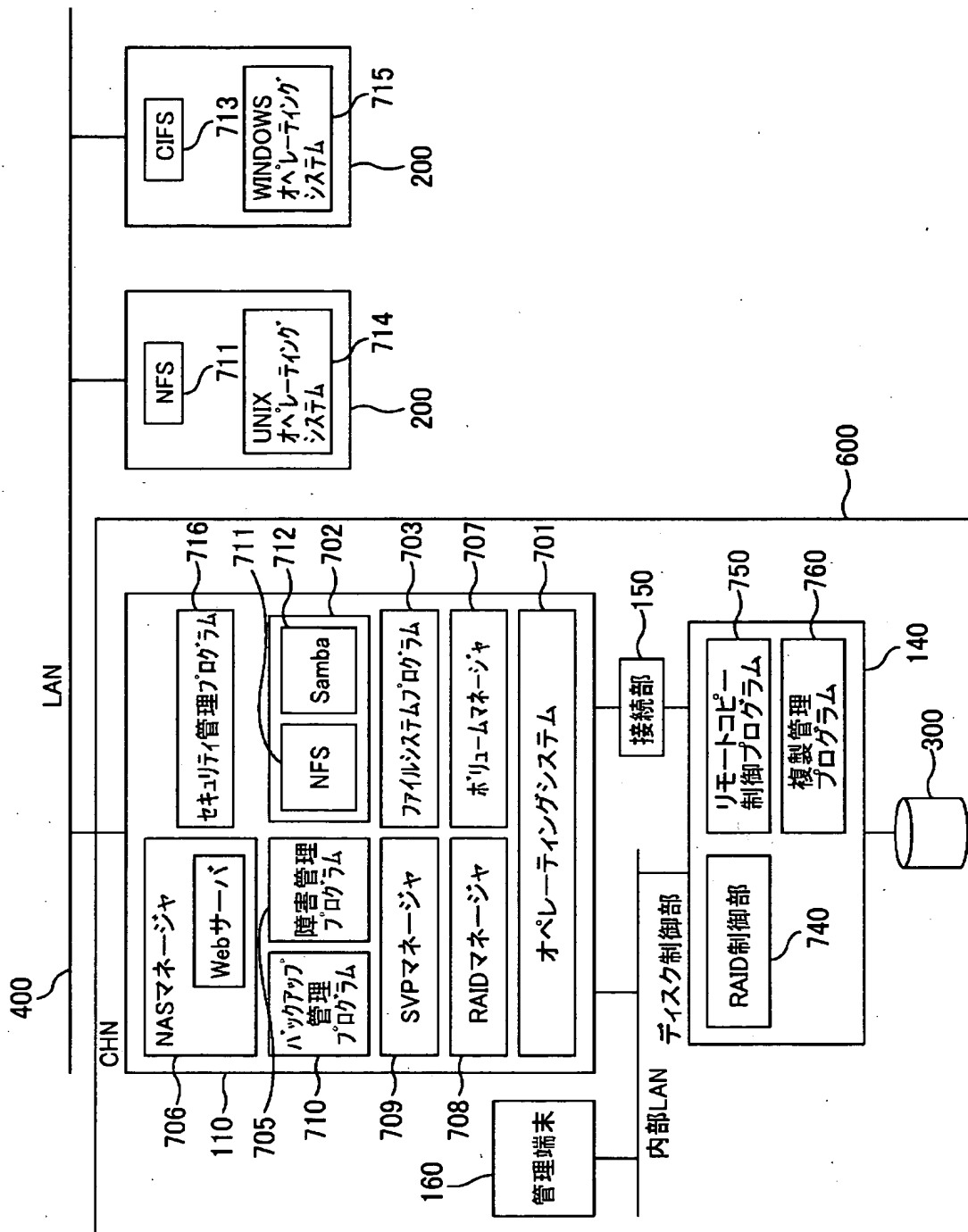
【図 9】



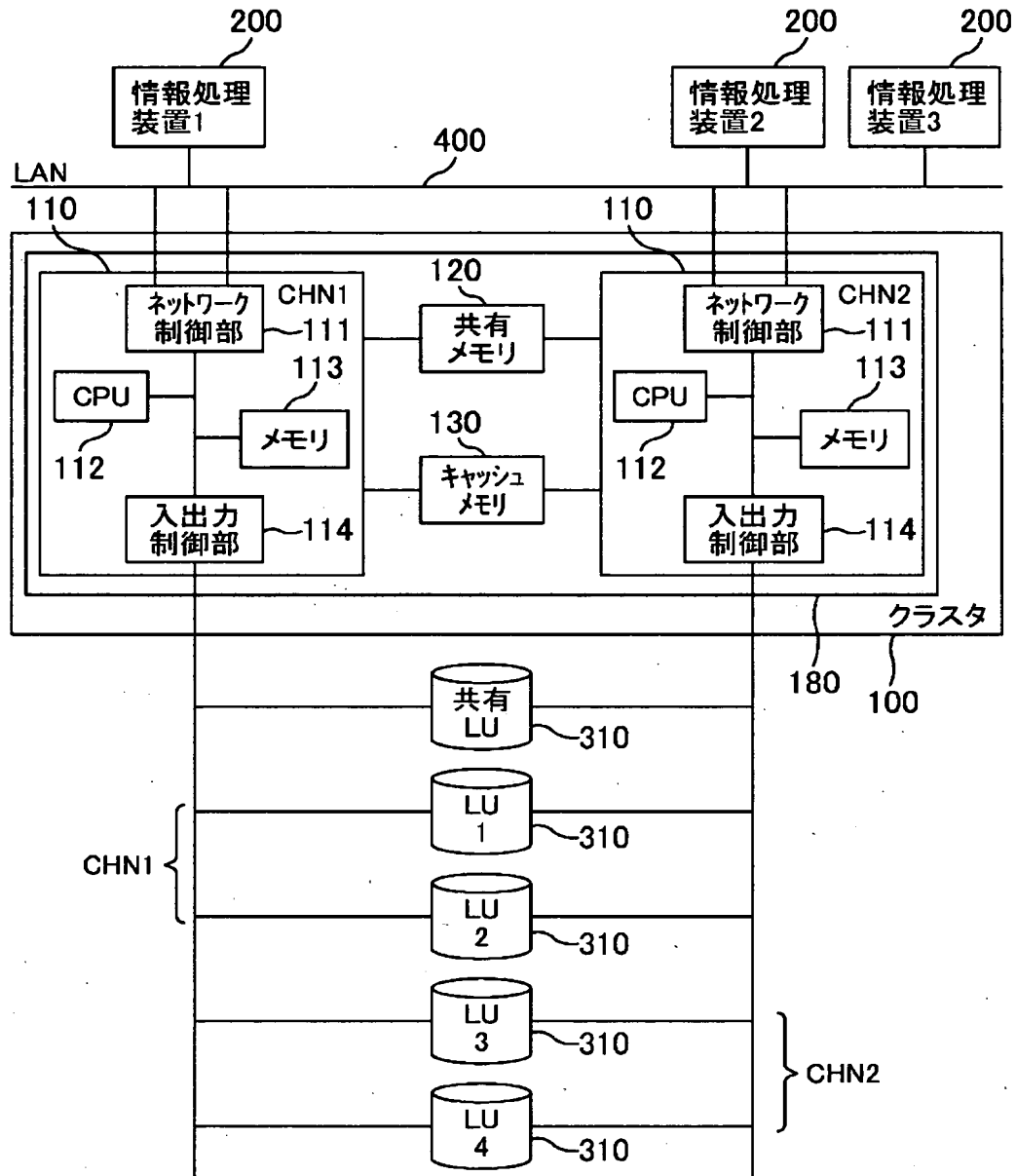
【図 1 0】



【図 11】



【図 1 2】



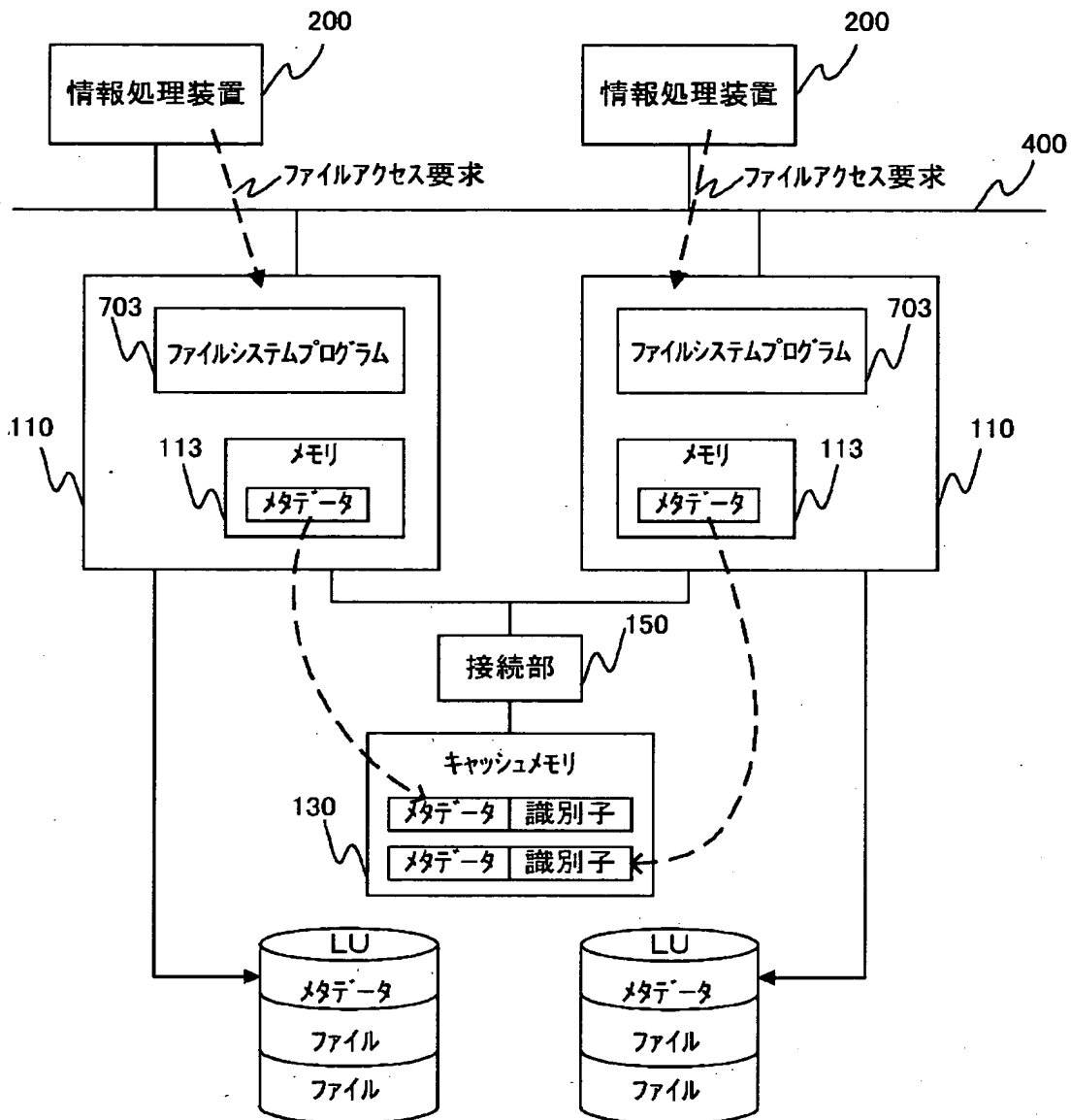
【図 1 3】

730
↘

メタデータ

ファイル名	先頭アドレス	容量	所有者	更新時刻
A	7BSA	200MB	X	0:00
B	05BF	50MB	X	7:57
C	1F30	100MB	Y	9:15
D	470B	100MB	Z	15:20
⋮	⋮	⋮	⋮	⋮

【図 1 4】



【図 1 5】

721

ファイルロックテーブル

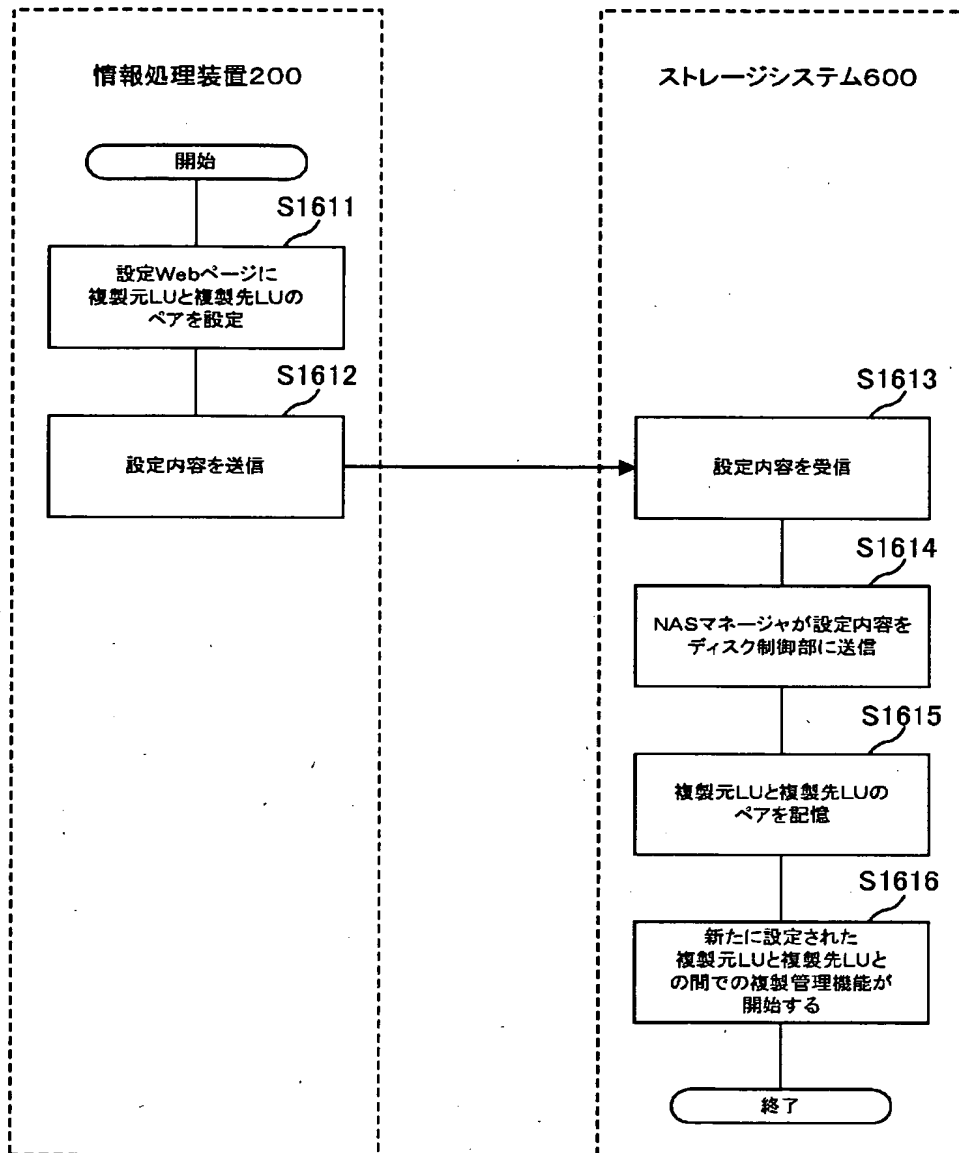
ファイル名	ロック状態
A	ロック中
B	—
C	—
D	ロック中
⋮	⋮

722

LUロックテーブル

LU	ロック状態
共有	—
1	ロック中
2	—
⋮	⋮

【図16】



【図 17】

1700

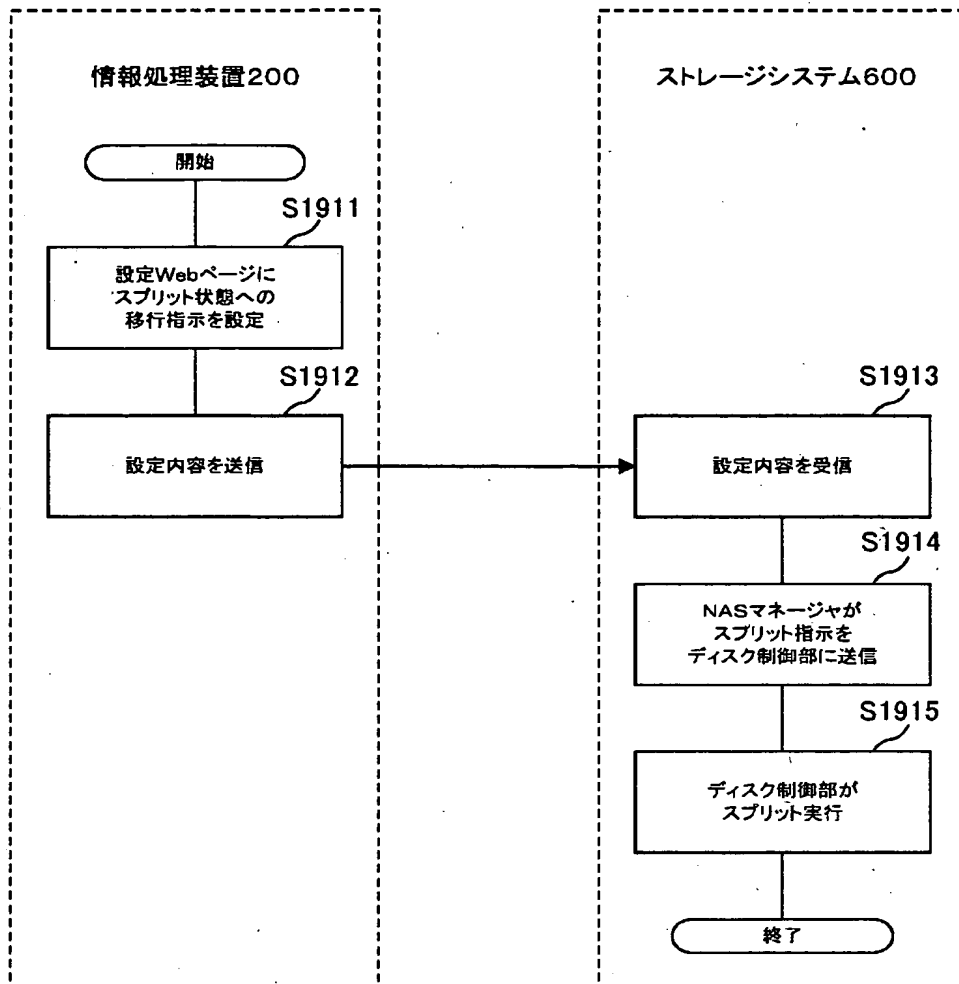
複製元LU (LUN)	複製先LU (LUN)
1	10
2	11
3	12
4	13
5	14

【図 18】

1800

複製元LU (LUN)	複製先LU (LUN)	スプリット	
1	10	実行	▽
2	11		▽
3	12		▽
4	13		▽
5	14		▽

【図 19】

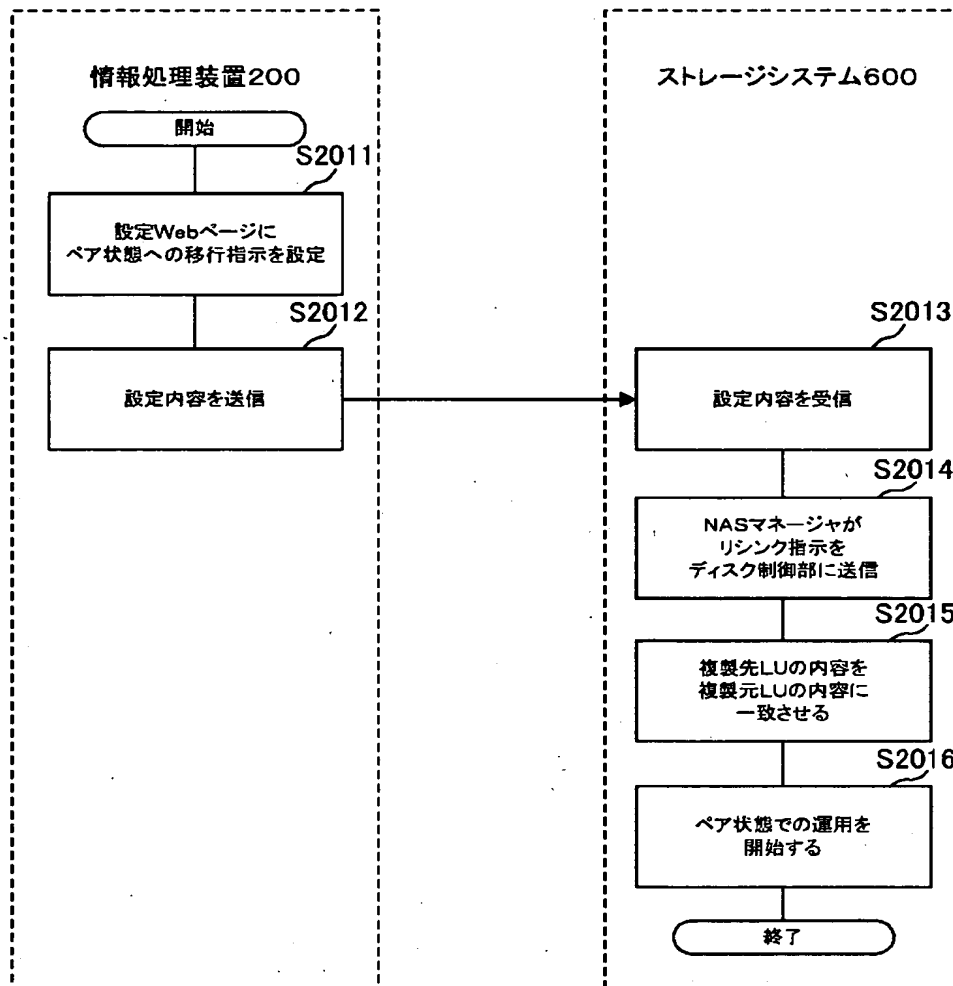


【図 20】

2000

複製元LU (LUN)	複製先LU (LUN)	リシンク
1	10	実行 ▾
2	11	▾
3	12	▾
4	13	▾
5	14	▾

【図 2 1】



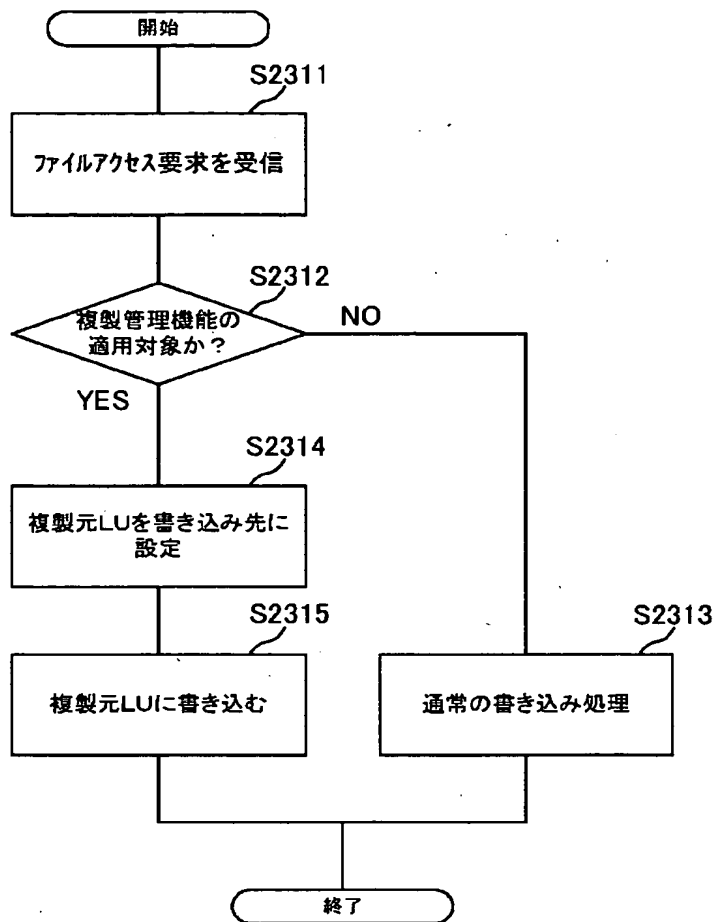
【図 2 2】

2200

ファイル名	ディレクトリ名
FILE1	DIR1
FILE2	DIR2
FILE3	
FILE4	

OK
キャンセル

【図 23】



【図 24】

Figure 24 shows a dialog box (2400) for file management. It contains two tables and two buttons.

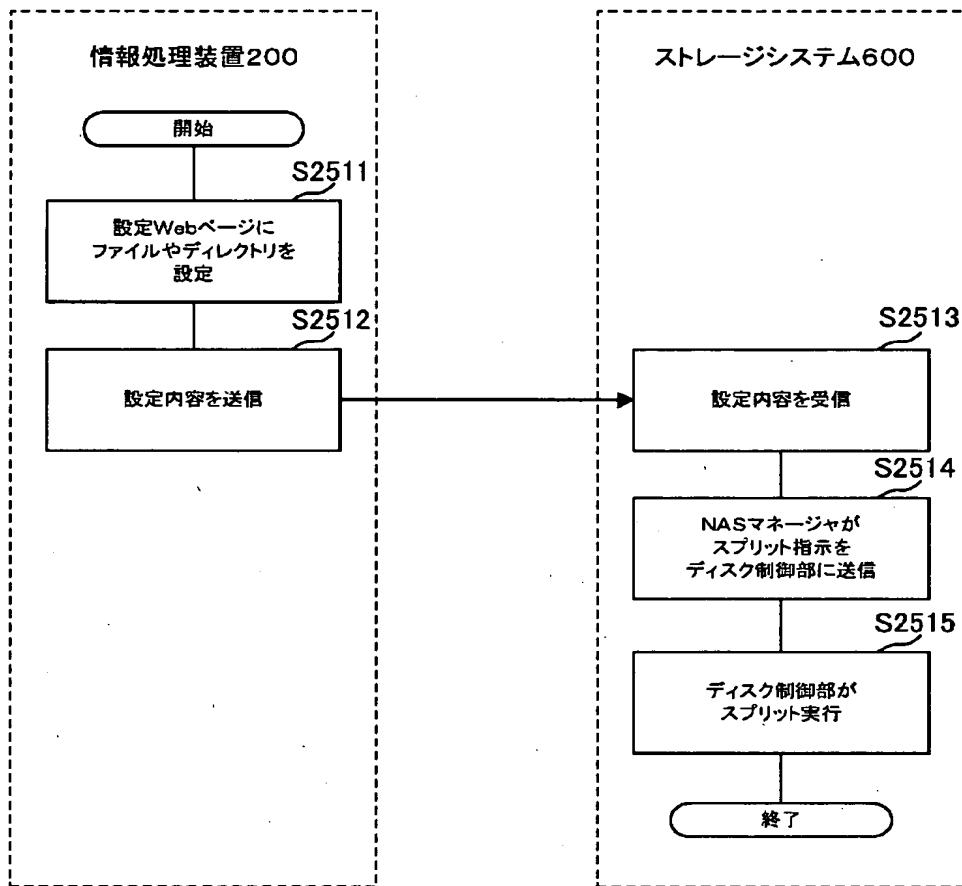
ファイル名	ディレクトリ名
FILE1	DIR1
FILE2	DIR2
FILE3	
FILE4	

日時指定		
	2003/1/30	1:00
	2003/1/30	2:00

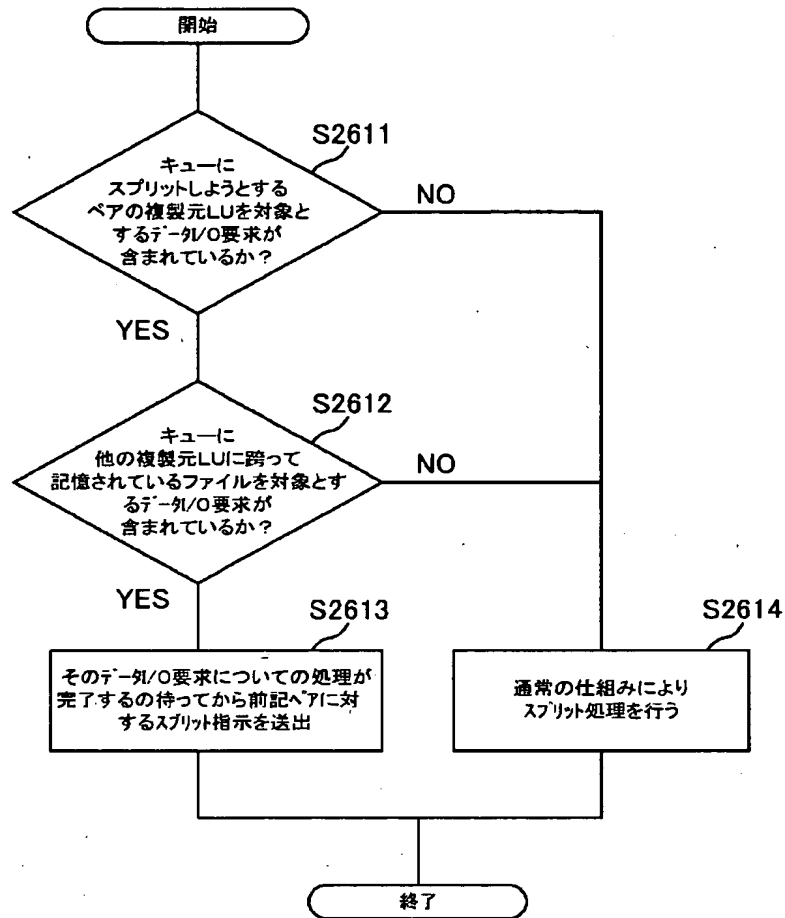
Buttons: **OK** **キャンセル**

Reference number: 2400

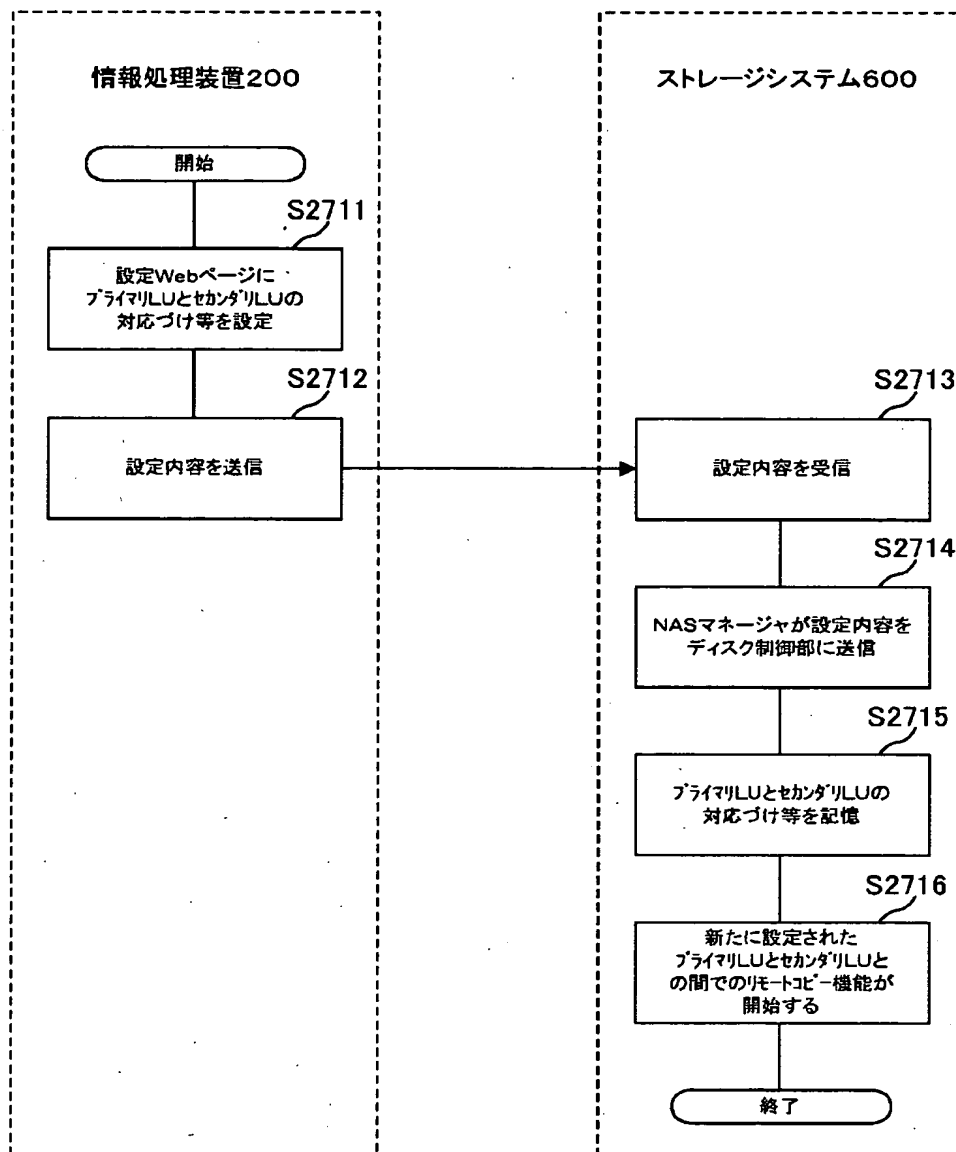
【図 25】



【図 26】



【図 27】



【図 2 8】

2800

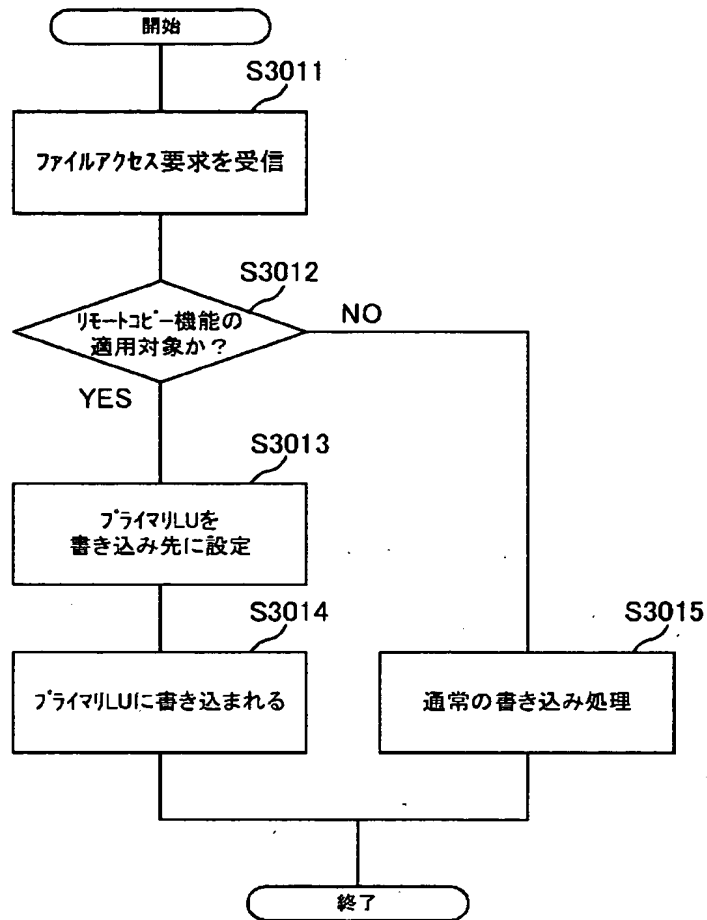
プライマリLU (LUN)	セカンダリLU (LUN)	方式	
1	10	同期	▽
2	11	非同期	▽
3	12	非同期	▽
4	13	非同期	▽
5	14	同期	▽

【図 2 9】

2900

ファイル名	ディレクトリ名
FILE1	DIR1
FILE2	DIR2
FILE3	
FILE4	

【図 3 0】



【書類名】 要約書

【要約】

【解決手段】 情報処理装置からファイル単位でのデータ入出力要求を受信するファイルアクセス処理部と、記憶デバイスに対するデータ入出力要求に対応するI/O要求を出力するI/Oプロセッサとが形成された回路基板を有する複数のチャンネル制御部と、I/Oプロセッサから送信される前記I/O要求に応じて記憶デバイスに対するデータ入出力を実行するディスク制御部とを含んで構成される記憶デバイス制御装置において、ディスク制御部が第一の論理ボリュームにデータが書き込まれた場合にそのデータの複製を第二の論理ボリュームにも記憶するために前記データを第二の論理ボリュームにも書き込む複製管理処理を実行するようにする。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地
氏 名 株式会社日立製作所